

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner #: _____ Date: _____
 Art Unit: _____ Phone Number 30 _____ Serial Number: _____
 Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

Point of Contact:
 Jan Delaval
 Librarian-Physical Sciences
 CM1 1E01 Tel: 308-4498

STAFF USE ONLY

	Type of Search	Vendors and cost where applicable
Searcher: <u>Jan</u>	NA Sequence (#) <u>✓</u>	STN _____
Searcher Phone #: <u>4498</u>	AA Sequence (#) <u>✓</u>	Dialog _____
Searcher Location: _____	Structure (#) _____	Questel/Orbit: _____
Date Searcher Picked Up: <u>9/14</u>	Bibliographic _____	Dr. Link _____
Date Completed: <u>9/12/01</u>	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: _____	Fulltext _____	Sequence Systems <u>✓</u>
Clerical Prep Time: <u>10</u>	Patent Family _____	WWW/Internet _____
Online Time: <u>15</u>	Other _____	Other (specify) _____

This Page Blank (copy)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 11, 2001, 11:38:28 : Search time 2570.69 Seconds
(without alignments)
11709.012 Million cell updates/sec

Title: US-09-192-611-1
Perfect score: 1946
Sequence: 1 ACAGTGTGGAGATGGCGGA.....CGCGATTCGCGCGCAGCTCGAG 1946

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_cm:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_rod:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_cm:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vl:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vl1:*
59: gb_vl2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	1946	100.0	1946	9	AR028527 Sequence
2	1946	100.0	1946	94	U76759 Mus musculu
3	741.4	38.1	2576	9	AX004160 Sequence
4	721.6	37.1	1260	9	AX004161 Sequence
5	721.6	37.1	3047	85	AB050511 Macaca fa
6	80	4.1	206008	76	AC084162 Mus muscu
7	80	4.1	210695	76	AC083753 Mus muscu
8	54.2	2.8	173126	86	AC005988 Homo sapi

9	51	2.6	224583	75	AC074153	
10	49.8	2.6	149719	78	AC090977	
11	49.2	2.5	214574	65	AC019272	
12	48.2	2.5	11696	3	SCPSCL	
13	47.4	2.4	160893	74	AC069456	
14	47.2	2.4	77836	67	AC022573	
15	47	2.4	147225	67	AC022498	
16	46.4	2.4	8438	59	HEPVIE	
17	46.4	2.4	8438	59	SHLIT	
18	46	2.4	482	7	AF251348	
19	46	2.4	249687	63	AC015693	
20	45.8	2.4	34862	3	SCG20A	
21	45.6	2.3	173561	64	AC016240	
22	45.4	2.3	16341	63	AC013381	
23	45.2	2.3	139627	64	AC016779	
24	45.2	2.3	160285	88	AC090945	
25	45.2	2.3	162403	23	AC016922	
26	45	2.3	533	54	G57145	
27	44.8	2.3	123580	2	AF263912	
28	44.8	2.3	152114	83	AP003448	
29	44.8	2.3	173154	76	AC079933	
30	44.6	2.3	51914	73	AC068022	
31	44	2.3	4242	9	AR106298	
32	44	2.3	5222	93	HSXDNA	
33	44	2.3	168013	78	AF261720	
34	44	2.3	209887	78	AC090649	
35	43.4	2.2	10022	1	AE005873	
36	43.2	2.2	588	89	AF251350	
37	43	2.2	123395	78	AC090433	
38	42.8	2.2	225257	74	AC069214	
39	42.6	2.2	144893	79	AL355486	
40	42.6	2.2	162345	90	AL161793	
41	42.6	2.2	186200	81	AL512306	
42	42.4	2.2	96299	86	AC066330	
43	42.2	2.2	20102	3	SC9B2	
44	42.2	2.2	69173	69	AC025052	
45	42	2.2	20622	94	AF230869	

ALIGNMENTS

RESULT 1						
AR028527	AR028527	1946 bp	DNA	PAT	29-SEP-1999	
LOCUS	Sequence	1 from patent US 5858711.				
DEFINITION	AR028527					
ACCESSION	AR028527.1	GI:5940500				
VERSION						
KEYWORDS						
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	1 (bases 1 to 1946)					
AUTHORS	Glincher, L.H. and Hodge, M.R.					
TITLE	NF-AT-Interacting protein NIP45 and methods of use therefor					
JOURNAL	Patent: US 5858711-A 1 12-JAN-1999;					
FEATURES	Location/Qualifiers					
source	1..1946					
BASE COUNT	438 a	500 c	562 g	446 t		
ORIGIN						

Query Match 100.0%: Score 1946; DB 9; Length 1946;
Best Local Similarity 100.0%: Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;
Matches 1946; Conservative 0;

QY	1	ACAGTGTGGAGATGGCGGAACCACTGAGGGACGTGGTCCGAGTCCCGGGTGGCCGA	60
DB	1	ACAGTGTGGAGATGGCGGAACCACTGAGGGACGTGGTCCGAGTCCCGGGTGGCCGA	60
QY	61	GCGGCTGGAGAGCCGAGGCGCCGCTGGCGGTCTCTGCGCCCGGCAAGTCTCGGGCT	120
DB	61	GCGGCTGGAGAGCCGAGGCGCCGCTGGCGGTCTCTGCGCCCGGCAAGTCTCGGGCT	120

DB	61	GCGGCTGGAGAGCCGAGGCGCCGCTGGCGGTCTCTGCGCCCGGCAAGTCTCGGGCT	120
QY	121	AGGCTATTCCAGACACCGGTGTTGTGAGCTGGTGTGATGACAGGACGAGAGAGTCTTG	180
DB	121	AGGCTATTCCAGACACCGGTGTTGTGAGCTGGTGTGATGACAGGACGAGAGAGTCTTG	180
QY	181	GAAGTGCAGAACCCAGTAGAGGTGCCGTGCGGCTCCCGCGCGCTAAACCTGAG	240
DB	181	GAAGTGCAGAACCCAGTAGAGGTGCCGTGCGGCTCCCGCGCGCTAAACCTGAG	240
QY	241	CAGAGAGGACAGTACAGTGAAGGGGGGCGGAGGGGCTCGGGAGCCCGGCTTACA	300
DB	241	CAGAGAGGACAGTACAGTGAAGGGGGGCGGAGGGGCTCGGGAGCCCGGCTTACA	300
QY	301	TGAGTGCAGAGCGGGGGGGGGGGTGTGATCCCGAGAGAGGCGCGGTGTCCAGTG	360
DB	301	TGAGTGCAGAGCGGGGGGGGGGGTGTGATCCCGAGAGAGGCGCGGTGTCCAGTG	360
QY	361	TACTCCGGAGGTACAGAGACGCTCAACTGATTCAGATTAATTCATTCCTTTGAAA	420
DB	361	TACTCCGGAGGTACAGAGACGCTCAACTGATTCAGATTAATTCATTCCTTTGAAA	420
QY	421	CTGTGCTTCAGAGCTGAAGATGAAGGAGATCTGACAAATTCGGAGTTCCCTCT	480
DB	421	CTGTGCTTCAGAGCTGAAGATGAAGGAGATCTGACAAATTCGGAGTTCCCTCT	480
QY	481	GAGAGATGATCCCTGCTTCAGGTTCTCCCTGAGAGAAAGCTCAGAAAGAGTGTAG	540
DB	481	GAGAGATGATCCCTGCTTCAGGTTCTCCCTGAGAGAAAGCTCAGAAAGAGTGTAG	540
QY	541	AAAGAGAAAGAAATGGAAGATTTCGGACACAGACATCTCTCTTTGCCCAACT	600
DB	541	AAAGAGAAAGAAATGGAAGATTTCGGACACAGACATCTCTCTTTGCCCAACT	600
QY	601	TGCTCAAGGAAAGAAAGCAATACGAGAGGCTCCAGAGTAAAGGAGAGAAC	660
DB	601	TGCTCAAGGAAAGAAAGCAATACGAGAGGCTCCAGAGTAAAGGAGAGAAC	660
QY	661	AAGCGTCTCCAAAGATCTCCGCTCTGAGGCCCAAGAGCAGCAGATCCAGCCCTT	720
DB	661	AAGCGTCTCCAAAGATCTCCGCTCTGAGGCCCAAGAGCAGCAGATCCAGCCCTT	720
QY	721	CAGAGCAGAGATGATGAGTGTCTAGTGAAGGAGCTGTCTTGGCCACAGACTTCGA	780
DB	721	CAGAGCAGAGATGATGAGTGTCTAGTGAAGGAGCTGTCTTGGCCACAGACTTCGA	780
QY	781	CTTTTACATCAAGATCCGCTGCGGGCTGACCTTGTAGACTGTGCTGAGATGTG	840
DB	781	CTTTTACATCAAGATCCGCTGCGGGCTGACCTTGTAGACTGTGCTGAGATGTG	840
QY	841	GAGCCCTTCAGAAATGTGTGATCAACATGAGCCCAATCAATCTTGGGTGTCTCAACAG	900
DB	841	GAGCCCTTCAGAAATGTGTGATCAACATGAGCCCAATCAATCTTGGGTGTCTCAACAG	900
QY	901	ATTCTTTTGGCTTTTGGAGAGTGAATCTCTACTGACCAACCCCTGATCACTTAAG	960
DB	901	ATTCTTTTGGCTTTTGGAGAGTGAATCTCTACTGACCAACCCCTGATCACTTAAG	960
QY	961	CTTGAGATGGCTACATCATGTGTTGTGTGCTGTAGCAAGCTCTTCAAGGCGCACAGAG	1020
DB	961	CTTGAGATGGCTACATCATGTGTTGTGTGCTGTAGCAAGCTCTTCAAGGCGCACAGAG	1020
QY	1021	ACATCCAGAGAGTCCGGGCTCCGGGTGCAAGGAGAGAGAACACCAAGATGTTGAGATC	1080
DB	1021	ACATCCAGAGAGTCCGGGCTCCGGGTGCAAGGAGAGAGAACACCAAGATGTTGAGATC	1080
QY	1081	TCAGTCTCTCTGATTTCTCTTTAAGGTTCTCATGTCAACACTATGAGAGAGCATGGA	1140
DB	1081	TCAGTCTCTCTGATTTCTCTTTAAGGTTCTCATGTCAACACTATGAGAGAGCATGGA	1140
QY	1141	CTCTCTGACACAAAGCTCTCTTTCTTTGATGGAGCAAAAGCTTTCAAGGAGAGCTG	1200
DB	1141	CTCTCTGACACAAAGCTCTCTTTCTTTGATGGAGCAAAAGCTTTCAAGGAGAGCTG	1200

Db	601	TCGTCAAGGAACAAACACAGAACGATACGAGAGCGCCTCCACAACCTAAGGAACTGAC	660
Qy	661	AAGCGTCTCCAAAGATCTCGCTCCTGCGCTGAGCCCCAGACAGCACAGAGTCCACGCTT	720
Db	661	AAGGCTCTCCAAAGATCTCGCTCCTGCGCTGAGCCCCAGACAGCACAGAGTCCACGCTT	720
Qy	721	CAGAGCACAGATGTATGTAGGTGTCCTAAGTGGAAAGGGCCGTCTTCCACAGAGCTCTCGA	780
Db	721	CAGAGCACAGATGTATGTAGGTGTCCTAAGTGGAAAGGGCCGTCTTCCACAGAGCTCTCGA	780
Qy	781	CTCTTTACACTCAAGATCCGGTCCCGGGCTGACCTTGTGAGACTGCCTGTGAGATGTCTG	840
Db	781	CTCTTTACACTCAAGATCCGGTCCCGGGCTGACCTTGTGAGACTGCCTGTGAGATGTCTG	840
Qy	841	GAGCCCTTCAGAAATGTGTGTGGATCACATGTGCCATCATCTTGGGGTGTCTTCAACACAG	900
Db	841	GAGCCCTTCAGAAATGTGTGTGGATCACATGTGCCAATCATCTTGGGGTGTCTTCAACACAG	900
Qy	901	ATTCCTTTGGCTTTTGGAGAGAGAACTGTCTCTACTGCCACCCCTAGTACCCCTAAG	960
Db	901	ATTCCTTTGGCTTTTGGAGAGAGATGAATCTCTCTACTGCCACCCCTAGTACCCCTAAG	960
Qy	961	CTTGGAGTGGCTGACATCATTTGATTTGTGTGGTCTAAGCAAGCTCTTCAGAGGCCACAG	1020
Db	961	CTTGGAGTGGCTGACATCATTTGATTTGTGTGGTCTAAGCAAGCTCTTCAGAGGCCACAG	1020
Qy	1021	ACATCCCAAGAGCTCGGGGCTCGGGGTGGAGGGGAAGGAACACACAGATGTGTAGATC	1080
Db	1021	ACATCCCAAGAGCTCGGGGCTCGGGGTGGAGGGGAAGGAACACACAGATGTGTAGATC	1080
Qy	1081	TCACGTCTCTCTATTTCTCCTTTAAGTGTCTATGTCACTATAGGAAGCCATGGGA	1140
Db	1081	TCACGTCTCTCTATTTCTCCTTTAAGTGTCTATGTCACTATAGGAAGCCATGGGA	1140
Qy	1141	CTCTCTGGACACAGCTCTCCTCTTCTTGTATGGGACAAAGCTTTCAGGCAAGAGCTG	1200
Db	1141	CTCTCTGGACACAGCTCTCCTCTTCTTGTATGGGAGCAAAAGCTTTCAGGCAAGAGCTG	1200
Qy	1201	CCAGCTGATCTGGGCTCGGAGATCCGGAGATCTCATGGAAGTGTGGGCTGGAAGCTCTAC	1260
Db	1201	CCAGCTGATCTGGGCTCGGAGATCCGGAGATCTCATGGAAGTGTGGGCTGGAAGCTCTAC	1260
Qy	1261	CCTGTTGGAGCGCAAGCCCAAGCATGTGGAGACATATGCTCCCAATTTATATGTGATTT	1320
Db	1261	CCTGTTGGAGCGCAAGCCCAAGCATGTGGAGACATATGCTCCCAATTTATATGTGATTT	1320
Qy	1321	TTTGGCCCCATTAAGGGCTTAACAGAACTGAATAGAACTGTGTACTTATTTATTTCTGG	1380
Db	1321	TTTGGCCCCATTAAGGGCTTAACAGAACTGAATAGAACTGTGTACTTATTTATTTCTGG	1380
Qy	1381	TGCTGGGGATTTGAACCCACAGACTATGCATGTCTAAGATGTATGAGTGGAGGCCAAAC	1440
Db	1381	TGCTGGGGATTTGAACCCACAGACTATGCATGTCTAAGATGTATGAGTGGAGGCCAAAC	1440
Qy	1441	CAAGGCATTACCTTTACCGAGCCCTAATAGATAGTATAGTCAACCACTGGCTACTCTGG	1500
Db	1441	CAAGGCATTACCTTTACCGAGCCCTAATAGATAGTATAGTCAACCACTGGCTACTCTGG	1500
Qy	1501	TAGTTGTGTGGCTGTGTATGTTTGTCTGTATTTTGGAGCCGCCCTGGGGCACATAGAG	1560
Db	1501	TAGTTGTGTGGCTGTGTATGTTTGTCTGTATTTTGGAGCCGCCCTGGGGCACATAGAG	1560
Qy	1561	GGACCTTGGCTTCCCTACCAATTCACGTTCCGCTGGTGGCCCTTCTCTATACAGATGACTT	1620
Db	1561	GGACCTTGGCTTCCCTACCAATTCACGTTCCGCTGGTGGCCCTTCTCTATACAGATGACTT	1620
Qy	1621	CTGTGAAGCTGCTATTTATAGTGTGTGTAACATAATAGAGCTCTCTTTGGGTGTCCAGG	1680
Db	1621	CTGTGAAGCTGCTATTTATAGTGTGTGTAACATAATAGAGCTCTCTTTGGGTGTCCAGG	1680
Qy	1681	CCTGGGGTTTGTGCCGAGTTTGGAGCCAGCAGTGAAGTCTACTGTCACTTGGAGCTGAGAA	1740
Db	1681	CTGTGGGGTTTGTGCCGAGTTTGGAGCCAGCAGTGAAGTCTACTGTCACTTGGAGCTGAGAA	1740

QY	1741	TGCATTTCCTGGTGGAGACACTCGGGGTGCAGAAATATAACAGAGTACATACATGCTG	1800
Db	1741	TCCATTTCCTGGTGGAGACACTCGGGGTGCAGAAATATAACAGAGTACATACATGCTG	1800
QY	1801	AAGCTGAGGACACTAGAGTCCGAAATGTAAACGACGCTTGGCATTTTCAAGCCTTGGGTATCCTCTCT	1860
Db	1801	AAGCTGAGGACACTAGAGTCCGAAATGTAAACGACGCTTGGCATTTTCAAGCCTTGGGTATCCTCTCT	1860
QY	1861	GCCCTGCCAGACACTAGACCAAGTGTGTGTACACACTTCTTGGCATGACACACTAGCTGGA	1920
Db	1861	GCCCTGCCAGACACTAGACCAAGTGTGTGTGTACACACTTCTTGGCATGACACACTAGCTGGA	1920
QY	1921	CGCGGGCGCGGATTCGGCCGACCTCGAG	1946
Db	1921	CGCGGGCGCGGATTCGGCCGACCTCGAG	1946
RESULT	3		
LOCUS	AX004160	2576 bp	DNA
DEFINITION	Sequence 1 from Patent WO921993.		
ACCESSION	AX004160		
VERSION	AX004160.1	GI:9927712	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 2576)		
TITLE	Zhao,J., and Zhou,H.		
JOURNAL	Human nlp-45: Interleukin-4 gene transcriptional trans-activator and its uses		
FEATURES	Patent: WO 921993-A 1 06-MAY-1999; ZENCA LTD (GB)		
source	Location/Qualifiers		
	1..2576		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
BASE COUNT	549 a 690 c 709 g 628 t		
ORIGIN			
Query Match	38.1%;	Score 741.4;	DB 9;
Best Local Similarity	74.5%;	Pred. No. 9.4e-168;	
Matches 1033;	Conservative	0;	Mismatches 306; Indels 48; Gaps
QY	6	GCGGAGATGCGCGGAAACACTGAGGGGAGAGTGTCCAGGTCCCGCGT-----GG	56
Db	6	GTCGTGCATGCGGGAGCGCCTGTGGGGAGCGGGCGCGGTGTCGGAGGTAGCGGTCCGG	65
QY	57	CCGAGGCGCTCGGAGAGCCCGAGGCGCCCGTGGCGGTGTCTCGCGCCCGGCACTTCC	116
Db	66	CCGAGGCGGTCCGGGCGCGGTGGGGGTCGGCGCGCGCTCTTCGGGCCCAACGGCTTCC	125
QY	117	GCGTGAAGCTCATTCAGACACCGCTGCTGTGTGACTTGGTCATGTGACAGGAGAGAGT	176
Db	126	ATTCGCGGCGCACTGACGTAGTGTCTGTGTGACTTGGTCACCGCAGCGAGTAGAGAAAT	185
QY	177	CTTGAAGT-----CGCAGACCCAGTAGAGGTGCCGTCGCCGCT	218
Db	186	TCTGAGAGTCCGACCGGTCGCGGTGCCCGGACGAGGTTCGAGTGTGAGACCCCGAGACC	245
QY	219	CCCCGCGCGGTAAACTGAGACAGACAGCAACGTACACTGAAGGGGCGCGCGAGG	278
Db	246	CCCGGCGCGGTGCGGTCCCGGATTAACAGCAACAGTACACAGCAAGGAGGAGAGGCG	305
QY	279	GCGTCGAGGAGCCCGCGCTACATTGTGTGGAGAGCGGGCGCGCGGTGTGATCCCGG	338
Db	306	GCGCGGAGACCCCGCGCGGAGCGGTACAGGCGGCGCGGGGTGTGTGTGTGTGTGTGT	365
QY	339	AGAGCGCGCGGTGTCCAGTGTACTCCGGGAAGGTACAGAGCAGCCCTCAACTTCATTC	398

Db 366 GGAGGGCCGCTGGTTCGGGTGACTCGGGAGGTTAAAGACGCTTCCGCTTATCCC 425
QY 399 AGATATATCATCCCTCTTGAACCTGTGCTTCAAGAGCTTGAAGATGAGGACGATCTGAC 458
Db 426 AGATGATCTATCCCTCGAAGACTCTACCTCGAGGGATGAGAGAGGACGAGACTGGC 485
QY 459 AAATTTGGGAGTTCCTCTGAGAGATGATCCCTGCTTCAAGTCTCTCCCTGAGAAA 518
Db 486 AGATTC---GAGTGGTCTCTACCATGAGGGCTCCCATCAGAGCTCTCCCTGAGAAC 542
QY 519 GAAGCTGAGAGAGAGTGTGAGAAAGAAAGAAATGAGAGGTTCCGGACGACGAGA 578
Db 543 AAAGCTGAGGACTAG---GATAAAGAGAGAGAAAGAAAGAGGTTCTGAGATCTGA 599
QY 579 CATCTCTCTTGGCCCAACCTTGTCAAGAGAACAAAGAGAGATGAGGAGGAGCT 638
Db 600 CAATCTCTCTGTGCTCCACCTTCAACGAGAGCAAAAGAGAGGATCTACTCGGAGCT 659
QY 639 CCAGAGCTAGGAGAGTGAACAAGCGTCTCCAGATCTCGGCTCTGAGAGCCGAA 698
Db 660 CAAGAGTTAGTGAAGTGAACAAGCGCTCCAGAGATCTCGGTTCTGTGAGAGCCCA 719
QY 699 GCAGACCAAGAGTCCAGCCCTTCAAGACAGATGATGATGCTCTGATGAGAGGCC 758
Db 720 GCCACCTCAGGAGTGAAGCAACAGGAGGAGAGATGATGATGCTGTTGAGAGGAG 779
QY 759 TGTCTGGCAGAGAGCTCTGACTCTTACACTGATGATGCTGGGCTGGGCTGAGCTAGT 818
Db 780 CACCTCTCCAG 839
QY 819 GAGAGCTCTCTGAGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 878
Db 840 CAGATGAGGCTCTGAGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 899
QY 879 TCTTGGGCTGTCTCAAGAGAGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 938
Db 900 CTTGGGGGCTCTCCAGAGAGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 959
QY 939 TGCCACCCCTGATACCTTAAAGCTTGAAGTGTGATGATGATGATGATGATGATGATG 998
Db 960 TGCCACCTCCAG 1019
QY 999 AAGCTCTTCAAG 1058
Db 1020 AAGTCTCTCAAG 1079
QY 1059 GAACACCAAGATGTTGAGAGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1118
Db 1080 GAACACCAAGATGTTGAGAGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1139
QY 1119 ACACATGAG 1178
Db 1140 CCACATGAG 1199
QY 1179 AAAGCTTCAAG 1238
Db 1200 AAAGCTTCAAG 1259
QY 1239 AGTCTGGGAGTGA---AGCTCTACCTCTGTTGAGAGAGAGAGAGAGAGAGAGAG 1297
Db 1260 GGTCTGGGAGTGA---AGCTCTACCTCTGTTGAGAGAGAGAGAGAGAGAGAGAGAG 1319
QY 1298 CTCCCAATTTATATGATTTTTCGCGCATTAAGGCTTAACGAGATGATTAAGAA 1357
Db 1320 C-----TTTCCCTTTTTCGCGCATTAAGGCTTAAGGCTTAAGGCTTAAGAA 1365
QY 1358 CTGTTT 1364
Db 1366 CTATCT 1372

RESULT 4
AX004161

LOCUS AX004161 1260 bp DNA
DEFINITION Sequence 2 from Patent W09921993.
ACCESSION AX004161
VERSION AX004161.1 GI:9927713
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1260)
AUTHORS Zhao, J. and Zhou, H.
TITLE Human nlp-45: Interleukin-4 gene transcriptional trans-activator
and its uses
JOURNAL Patent: WO 9921993-A 2 06-MAY-1999;
ZENECA LTD (GB)
FEATURES
source location/Qualifiers
1..1260
BASE COUNT 269 a 364 c 407 g 220 t
ORIGIN
Query Match 37.1%; Score 721.6; DB 9; Length 1260;
Best Local Similarity 76.4%; Pred. No. 5.8e-163;
Matches 934; Conservative 0; Mismatches 264; Indels 24; Gaps 3;
QY 48 CCGCGGTGCGGAGGCGCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 107
Db 45 CCGTCCCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 104
QY 108 GCAGTCTCGGCTAGGCTCATTCAGACACCGTGTGAGCTTGTGACCTGTGACGACAGCGA 167
Db 105 GCGGTCTCATTCGCGGAG 164
QY 168 CGAAGAGTCTTGAAGT-----CGCAGACCGATGAGAGTCCCGGT 209
Db 165 TGAGGAATTTCTGAGAGTCCGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 224
QY 210 CGCCGCGCTCCCGCGGCGGCTAAACCTGAGACAGAGAGAGAGAGAGAGAGAGAGAG 269
Db 225 CCGGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 284
QY 270 GAGCGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 329
Db 285 GAGCGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 344
QY 330 GGATCCGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 389
Db 345 GGATCCGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 404
QY 390 CCTATTCAGATATTTATCCCTTGAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 449
Db 405 CTTATTCAGATATTTATCCCTTGAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 464
QY 450 AGATCTGAACAATTTCTGAGATCTCTCTCTGAGAGATGATCCCTCTGAGATTTCTCC 509
Db 465 AGAGCTGGACAGTCT---GAGTGTCTTACCATGAGAGAGAGAGAGAGAGAGAGAGAG 521
QY 510 CTGGAAGAGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 569
Db 522 CTGGAAGAGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 578
QY 570 GGAGCAGAGATCTCTCTTGGCCCAACCTTGTGAGAGAGAGAGAGAGAGAGAGAGAG 629
Db 579 GGATCTGAGACACTCTCTCTGCTCCACCTTCAACAGAGAGAGAGAGAGAGAGAGAG 638
QY 630 GAGCGCGCTCCAGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 689
Db 639 TCGGCACTCAAGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 698
QY 690 GAGCCCAAG 749

```

Db      699 GAGCCCCAAGCCACCTCAGGGGTCAAGAGCAACAGGCCCAAGAGGATGAAGTGTCTTGGT 758
Oy      750 GGAAGGGCGCTGTCTTGGCACAGAGCTCTTACACTGAAGATCCGGTCCCGGGC 809
Db      759 GGAAGGGCCACCTCCACAGAGACCCCCGACTCTTCCCACTCAAAATCCGTTCCCGGGC 818
Oy      810 TGACCTAGTAGACTGCGCTGTGAGATGTGAGACCCCTTCAGAAATGTGTGATGATACAT 869
Db      819 TGACCTAGTAGATTTGGCCCTCAGAGATGTGAGACCCCTGAGAGTGTGGTGACCAAT 878
Oy      870 GGGCAATCATCTTGGGGTGTCTCCAAACAGAGATCTTTTGGCTTTTGGAGAGAGTGAAT 929
Db      879 GGGCACCCACCTTGGGGTGTCTCCCAAGAGAGATCTTTTGGCTTTTGGAGAGAGAGCT 938
Oy      930 GTCTCTACTGCGCACCCCTAGTATACCTTAAAGCTGAGATGGCTGACATCATTTGATGTGT 989
Db      939 ATCAGCTACTGCGCACCTCCAGAGACCTTAAAGCTGGAGATGGCTGACATCATTTGATGTGT 998
Oy      990 GGTGCTAGCAAGCTCTTTCAGAGGCCACAGAGACATTCACAGAGCTCCGGCTCCGGGTGCA 1049
Db      999 GGTACTAAAGTCTTCCAGAGGCCACAGAGACGTCCCAAGAGCTCCAGAGCTCCGGGTGCA 1058
Oy      1050 GGGGAAGAGAAACACAGAGATGTGAGATGTCTGCTCTCTCTGATCTCTCTTAAGGT 1109
Db      1059 GGGAAAGAGAAACACAGAGACTGGAAGTCTCTGATCTCTCTGATCTCTCTTAAGAGC 1118
Oy      1110 TCTCATGTCACTAGTGAAGAGCCATGGACTCTCTGACACAGCTCTCTCTCTTCTT 1169
Db      1119 CCTCATGTCCACTATGAGAGGCCATGGACTCTCTGAGAGAGAGCTCTCTCTCTTCTT 1178
Oy      1170 TGATGGACAAGCTTTTCAGCAAGAGAGCTGCCAGCTGATCTGGGCTCGAATCCGGAGA 1229
Db      1179 TGATGGACAAGAGCTTTTCAGCAAGAGAGCTGCCAGCTGATCTGGGCTCGAATCTGGGGA 1238
Oy      1230 TCTCATGAAGTCTGGGGCTGA 1251
Db      1239 CCTCATTTGAGGCTTGGGGCTGA 1260

RESULT 5
AB050511 3047 bp mRNA PRI 27-OCT-2000
LOCUS Macaca fascicularis brain cDNA, clone:Ompa-16750.
DEFINITION AB050511.1 GI:11041476
VERSION f1s (full insert sequence).
KEYWORDS Macaca fascicularis adult male brain parietal lobe cDNA to mRNA, clone_1lb:macaque brain cDNA library Ompa clone:Ompa-16750.
SOURCE Macaca fascicularis
ORGANISM Macaca fascicularis
REFERENCE 1 (sites)
AUTHORS Osada,N., Hida,M., Kusuda,J., Tanuma,R., Iseki,K., Hirai,M., Terao,K., Suzuki,Y., Sugano,S. and Hashimoto,K.
TITLE Isolation of full-length cDNA clones from macaque brain cDNA libraries
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3047)
AUTHORS Hashimoto,K., Osada,N., Hida,M., Kusuda,J. and Sugano,S.
TITLE Direct Submission
JOURNAL Submitted (25-OCT-2000) to the DDBJ/EMBL/Genbank databases.
Katsuyuki Hashimoto, National Institute of Infectious Diseases, Division of Genetic Resources; 23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan (E-mail:khashi@nii.go.jp, URL:http://www.nii.go.jp/yoken/genbank/, Tel:81-3-5285-1111(ex.2120), Fax:81-3-5285-1181)
Lab host: TOP10
Vector: PME18S-FL3 (Acc.No. AB009864)
R. Site1: DraIII (CACTGTG)
R. Site2: DraIII (CACCATGC)
Description: 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCCTTTTCTTTTCTTTT]; double-stranded cDNA was synthesized
```

```

using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5kb. The SfiI-digested PCR product was cloned into distinct DraIII sites of pME18S-FL3. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Libraries were constructed by Sugano et al. (University of Tokyo, Institute of Medical Science). Custom primer used for sequencing (5' end primer [CTTGTGCTTAAAGCTCCG]; 3' end primer [CGAGCTCAGCTCAGCAGCA]).
FEATURES
    source
        1..3047
            /organism="Macaca fascicularis"
            /db_xref="taxon:9541"
            /clone="Ompa-16750"
            /sex="male"
            /tissue.type="brain parietal lobe"
            /clone.lib="macaque brain cDNA library Ompa"
            /dev_stage="adult"
            6..1232
            /codon_start=1
            /product="hypothetical protein"
            /protein_id="BAB17279.1"
            /db_xref="GI:11041477"
            /translation="MAEPVGRKRLSRSGAGRRPAQSPRSKGTIDVYSVDLVSND
            EETVEVTSACADEVEVAPSEPPGVASRDSDSBEADAPAPPEPRRRRL
            VLDGPAPLVPVYSKRVKSLCLIPDLISLKLKPPGDEEVELADSLYHGSFSP
            GSPWKTRLKRLDKERKRTSLDLSDSPISPSPRKSRKRLKLSLVKRLDPL
            RSLDSPEDPGOGDEDEVYVECPPTLPEFRPLPLKIRCADLVRLPLNRSPLD
            SYVDHMAHTHGVSPSRLLILFGFETSLPATPRYLKLVADIIDCVLASPEAPETS
            RQQLRVQGRKREKQTLFVLSLSRSPSLKTLMSHTEEMGLSGKRLSFFPGTSLSGREL
            PADLMEBSGLIEVWG"
BASE COUNT 676 a 797 c 821 g 753 t
ORIGIN
Query Match 37.1%; Score 721.6; DB 85; Length 3047;
Best Local Similarity 75.5%; Pred. No. 5.3e-163;
Matches 978; Conservative 0; Mismatches 279; Indels 39; Gaps 5;
Oy      88 GGGCGGTGCTCGCGCCGCGGACAGTCCGGCTAGGCTCAATTCAGACACGTCCTGTG 147
Db      57 GGGCGGCGTCTCTCGGGCCCGACGGGTCTTCATCCGGGGACACTGAGCGTGTCTGTG 116
Oy      148 GACTTGTCAAGTGAACGACGAGGAGAGGTCTTGAAGT-----CGCA 189
Db      117 GACTTGTCAAGCAGACAGCATGAGAAATGTGAGAGTCAACAGTCTCGCTGCGCGC 176
Oy      190 GACCACTAGAGTGGCGGCTGCCCGCTCCCGCGCTAAGCTGAGCAGAGACG 249
Db      177 GACGAGGTTGAGGTTGACACCTCGAGACCCCGGGCGGCTCGCGTCCGGATGACAGC 236
Oy      250 GACAGTGACAGTGAAGGCGCGCGCGAGGCGCTCGGAGCGCCGCTACATTGTGTCGA 309
Db      237 GACAGTACACGAGGAGGCGGAGCGCGCGCGCGACAGACCCCGCGGAGCGGATAGG 296
Oy      310 CGGCGGCGGCGGCGGCTGTGATCCCGAGAGGCGCGGCTGTGCTCCAGTGTACTCGG 369
Db      297 CGGCGGCGGCGGCGGCTGTGATCCCGGAGAGCGCGCGCTCGGTGCTGACTCGG 356
Oy      370 AAGTACAGAGACGCTCAACCTCATTTCCAGATATATTCATCCCTTTGAACATGTGCCCT 429
Db      357 AAGGTAAAGTAAAGTACCTTTCCTATATCCAGATGACCTATCCCTCGGAACCTTACCT 416
Oy      430 TCAGAGCTGAAGTGAAGGAGATGTACAAATCTGTGGAGTCTCCCTCGAGAGATGAT 489
Db      417 CCAGGAGTGAAGGAGGAGGAGATGTGCGAGATTGCG---AGTGTCTCTACATGAGGGC 473
Oy      490 GCCCTGCTTCAAGTGTCTCCCTGGAGAAAGAGCTCGAAGAAAGTGTGAGAAAGAGAA 549
Db      474 TCAGCATCGCAGGCTCTCCCTGGAGAACAGCAAGCTGAGGACTAAG---GATPAAAGAGAG 530
Oy      550 AAGAAATGAGAGATTTCGGAGCAGACATCTCTCTTTGCCCAACCTTGCTGTAAGG 609
        1111111 1 1111 111 1 1111 111111 1 1111111 1 111
```

```

Db      531 AAGAAAAAGACAGAGATTCGTGATCTGGACAACTCTCTGTGCCCCACCTTCACCGAGG 550
Oy      610 AACAAAAGCAGAAAGCATACGAGCGCTCCAGACGTAAGGAGGAGGAACAAGCCGCTC 669
Db      591 ACCAAAAGCAGAAAGCATACGAGCGCTCCAGACGTAAGGAGGAGGAACAAGCCGCTC 650
Oy      670 CAAGATCTCCGCTCTCTGCTGAGCCCCCAAGCAGACAGCAGAGCCCTTCAGAGCACA 729
Db      651 CAGATCTCCGCTCTCTGCTGAGCCCCCAAGCAGACGTAAGGAGGAGGAACAAGCCGCTC 710
Oy      730 GATGATGAGGTGTCTCTAGTGGAGAGGCTGTCTTGGCCACAGAGCTCTGACTCTTTACA 789
Db      711 GAGGAGGAAGTGTCTCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 770
Oy      790 CTCGAATGCGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 849
Db      771 CTCGAATGCGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 830
Oy      850 CAGAATGAGTGATGATCATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 909
Db      831 CAGAGTGTGTGAGACACATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 890
Oy      910 CTTTGGAGAGAGTGAATGATCTCTACTGACCCCTAGTACCTTAAAGCTTGGAGTG 969
Db      891 CTTTGGAGAGAGTGAATGATCTCTACTGACCCCTAGTACCTTAAAGCTTGGAGTG 950
Oy      970 GCTGATCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1029
Db      951 GCTGATCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1010
Oy      1030 GAGCTCCGAGCTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1089
Db      1011 CAGCTCCAGCTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1070
Oy      1090 CCGATTCCTCTCTTAAAGGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1149
Db      1071 CCGATTCCTCTCTTAAAGGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1130
Oy      1150 CACAAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1209
Db      1131 CGGAAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1190
Oy      1210 CTGGGCTCTGGAATCGGAGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1268
Db      1191 CTGGGCTCTGGAATCGGAGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1250
Oy      1269 GACGCAAGCAAGACATGAGAGCATAGTCCCAATTTATATGATTTTGGCCC 1328
Db      1251 ATGGCCACGCTGAGCTGGGAGAAATGACTCCC-----TGTTTTGGCC 1296
Oy      1329 CATTAAGGCTTAAAGCAAGTGAATTAAGTGAATTTGTTT 1364
Db      1297 CATTAAGGCTTAAAGCAAGTGAATTAAGTGAATTTGTTT 1332

RESULT 6
AC084162 206008 bp DNA HTG 07-FEB-2001
LOCUS Mus musculus chromosome 5 clone RP23-11P12 strain C57BL6/J, WORKING
DEFINITION DRAFT SEQUENCE, 14 unordered pieces.
ACCESSION AC084162 GI:12704629
VERSION AC084162.2 GI:12704629
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 206008)
REFERENCE
AUTHORS
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Ayale, R., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W.,
Buffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S.,
Guan, X., Gupta, J., Ho, S.-L., Idol, J.R., Karlins, E., Lee-Lin, S.-Q.,
Legsmpi, R., Lim, M., Maduro, O.L., Maduro, V.B., Mastaglio, C.,
Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Prasad, A.,

```

Shevchenko, Y., Snyder, B., Stantrijop, S., Thomas, J.M., Thomas, P.J.,
 Tjongson, E.E., Touchman, J.W., Tsugeon, C., Vogt, J.L., Walker, M.A.,
 Wetherby, K.D., Zhang, L.-H. and Green, E.D.
 NISC Comparative Sequencing Initiative
 Unpublished
 2 (bases 1 to 206008)
 Green, E.D.
 Direct Submission
 Submitted (14-OCT-2000) NIH Intramural Sequencing Center, 8717
 Givemont Circle, Gaithersburg, MD 20877, USA
 On Feb 7, 2001 this sequence version replaced gi:10801352.

COMMENT

Center: NIH Intramural Sequencing Center
 Center code: NISC
 Web site: <http://www.nisc.nih.gov>
 Contact: nisc_mouse@nigri.nih.gov
 ----- Project Information -----
 Center project name: rs
 Center clone name: 011P12
 ----- Summary Statistics -----
 Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 20092 bases at least Q40
 Consensus quality: 20157 bases at least Q30
 Consensus quality: 20200 bases at least Q20
 Insert size: 20200; agarose-fp
 Insert size: 204708; sum-of-ctrls
 Quality coverage: 11.83x in Q20 bases; agarose-fp
 Quality coverage: 11.68x in Q20 bases; sum-of-ctrls

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 14 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

FEATURES

source

1.206008
 /organism="Mus musculus"
 /strain="C57BL6/J"
 /db_xref="taxon:10090"
 /chromosome="5"
 /clone="RP23-11P12"

```

misc_feature      /clone_lib="RPCi mouse BAC library 23"
                  1..2503
misc_feature      /note="assembly_fragment"
                  2604..6058
misc_feature      /note="assembly_fragment"
                  6159..12381
misc_feature      /note="assembly_fragment"
                  12482..17111
misc_feature      /note="assembly_fragment"
                  clone_end:SP6
                  vector_side:left"
misc_feature      17212..24466
misc_feature      /note="assembly_fragment"
                  24567..33228
misc_feature      /note="assembly_fragment"
                  33329..40894
misc_feature      /note="assembly_fragment"
                  40995..52203
misc_feature      /note="assembly_fragment"
                  52304..64564
misc_feature      /note="assembly_fragment"
                  64665..76971
misc_feature      /note="assembly_fragment"
                  77072..92317
misc_feature      /note="assembly_fragment"
                  92418..113190
misc_feature      /note="assembly_fragment"
                  113291..159328
misc_feature      /note="assembly_fragment"
                  159429..206008
misc_feature      /note="assembly_fragment"
                  206009..51451
BASE COUNT      51690 a 49057 c 51451 g 52506 t 1304 others
ORIGIN

```

```

Query Match      4.13: Score 80; DB 76; Length 206008;
Best Local Similarity 73.13; Pred. No. 7.1e-09;
Matches 117; Conservative 0; Mismatches 40; Indels 3; Gaps 1;

```

```

OY 903 TCTTTCGCTTTTGGAGAGTGAAGTGTCTCTGACCCGCTGATACCTTAAGCT 962
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 156786 TTTTTCGCTTTTGGAGAGTGTGACCTGTCCCTTGTGCCACCCCTGATTTCTTAACCT 156845
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 963 TGAAGTGCCTGACATCATTTGATGTGTGTGTAGCAAGCTCTTCAGAGGCCACAGAGAC 1022
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 156846 TGTGTAGCGGACATCATCTGTTTG---CTGCTAGTAAAGCGCTTCCAAAGCTACTGAGAC 156902
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 1023 ATCCGAGAGCTCCGCGCTCCGCGTGCAGGAGGAGAGAGAAA 1062
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 156903 ATCCGAGAGTCCGCGCTCCGCGGAGAGAGAGAGAAA 156942
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 7
LOCUS      AC083753/c
DEFINITION Mus musculus chromosome 5 clone RP23-480K9 strain C57BL6/J, WORKING
ACCESSION  AC083753
VERSION     AC083753.2 GI:13027533
KEYWORDS
SOURCE      HTGS: HTGS_PHASE1; HTGS_DRAFT.
ORGANISM    Mus musculus
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 210695)
AUTHORS    Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
            Bouffard,G.G., Brinkley,C., Brooks,S., Dietrich,N.L., Granlie,S.,
            Gupta,J., Ho,S.-L., Idol,J.R., Karlins,E., Lee-Lin,S.-O.,
            Legaspi,R., Lim,M., Maduro,Q.L., Maduro,V.B., Maselli,C.,
            Mestrlian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A.,
            Shevchenko,Y., Snyder,B., Stantrop,S., Thomas,J.W., Thomas,P.J.,
            Thompson,E.E., Touchman,J.W., Tsurenko,C., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D., Zhang,J., Zhang,L., H. and Green,E.D.

```

```

TITLE      NISC Comparative Sequencing Initiative
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 210695)
AUTHORS    Green,E.D.
TITLE      Direct Submission
JOURNAL    Submitted (30-SEP-2000) NIH Intramural Sequencing Center, 8717
            Groveomont Circle, Gathersburg, MD 20877, USA
            On Feb 21, 2001 this sequence version replaced g1:10440605.
COMMENT    -----
            Center: NIH Intramural Sequencing Center
            Center code: NISC
            Web site: http://www.nisc.nih.gov
            Contact: nisc-mouse@nih.gov
            Project Information
            -----
            Center project name: rt
            Center clone name: 480K09
            -----
            Summary Statistics
            Sequencing vector: plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 208188 bases at least Q40
            Consensus quality: 208941 bases at least Q30
            Consensus quality: 209303 bases at least Q20
            Insert size: 203000; agarose-fp
            Insert size: 209995; sum-of-ctrls
            Quality coverage: 11.96x in Q20 bases; agarose-fp
            Quality coverage: 11.56x in Q20 bases; sum-of-ctrls
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            * 1 2686: contig of 2686 bp in length
            * 2687 2786: gap of unknown length
            * 2787 8537: contig of 5751 bp in length
            * 8538 8637: gap of unknown length
            * 8638 17518: contig of 8881 bp in length
            * 17519 17618: gap of unknown length
            * 17619 33828: contig of 16210 bp in length
            * 33829 33928: gap of unknown length
            * 33929 74176: contig of 40248 bp in length
            * 74177 74276: gap of unknown length
            * 74277 114480: contig of 40204 bp in length
            * 114481 114580: gap of unknown length
            * 114581 151240: contig of 36650 bp in length
            * 151241 151340: gap of unknown length
            * 151341 210695: contig of 59355 bp in length.
            Location/Qualifiers
            1..210695
            /organism="Mus musculus"
            /strain="C57BL6/J"
            /db_xref="taxon:10090"
            /chromosome="5"
            /clone="RP23-480K9"
            /clone_lib="RPCi mouse BAC library 23"
            1..2686
            /note="assembly_fragment"
            2787..8537
            /note="assembly_fragment"
            8638..17518
            /note="assembly_fragment"
            clone_end:77
            vector_side:right"
            17619..33828
            /note="assembly_fragment"
            33929..74176
            /note="assembly_fragment"
            74277..114480
            /note="assembly_fragment"

```

misc_feature	114581..151240	/note="assembly_fragment"
misc_feature	151341..210695	/note="assembly_fragment clone_end:SP6 vector_side:left"
BASE COUNT	52952 a 51898 c 52183 g 52957 t	705 others
ORIGIN		
Query Match	4.1%: Score 80; DB 76; Length 210695;	
Best Local Similarity	73.1%: Pred. No. 7,1e-09;	
Matches 117; Conservative	0; Mismatches 40; Indels 3; Gaps 1;	
Oy 903	TCCTTTGCTTTTGGAGAGTGAAGTCTGCTCTACTGCCACCCCTAGTACCCTAAAGCT	962
Db 160754	TTTTTCGTTTGTGGAGAGGCTGAAGCTGCCCTTCTGCACACCCCTAGATCTCAACCT	160695
Oy 963	TGGAGTGGCTACATCATGTGATTTGTGTGTGCTGTACCAAGCTCTTTAGAGGCCACAGAGAC	1022
Db 160694	TGGGTGTACCGGACATCATCTGTTTG---CTGCTAAATGAAGGCTTCCCAAGGCTACTGAGAC	160638
Oy 1023	ATCCAGAGAGCTCCGGGCTCCGGGTCCAGGGGGAAGAGAA	1062
Db 160637	ATCCAGAGAGTCCCGAGCTCCGGGACAGAGAGAGAAAGAAACA	160598

RESULT	8
LOCUS	AC005988
DEFINITION	Homo sapiens chromosome 17, clone hRPK.299_G_24, complete sequence.
ACCESSION	AC005988
VERSION	AC005988.1
KEYWORDS	GI:4156128
SOURCE	htg.
ORGANISM	homo.
ORGANISM	Homo sapiens
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
ORGANISM	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 173126)
REFERENCE	Bliren,B., Linton,L., Nusbaum,C. and Lander,E.
REFERENCE	Homo sapiens chromosome 17, clone hRPK.299_G_24
REFERENCE	Unpublished
REFERENCE	2 (bases 1 to 173126)
REFERENCE	Bliren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
REFERENCE	Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,
REFERENCE	Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S.,
REFERENCE	Collamore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
REFERENCE	Donelan,L., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R.,
REFERENCE	Gage,D., Gardyna,S., Geraldery,K., Grant,G., Hagos,B., Heaford,A.,
REFERENCE	Herens,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
REFERENCE	Karats,A., Lehoczy,J., Macdonald,P., Marguis,N., McEwan,P.,
REFERENCE	McGurk,A., McKernan,K., Meldrum,J., Molla,M., Morris,W., Morrow,J.,
REFERENCE	Mychalacky,J., Nait,R., Naylor,J., Niloff,M., O'Connor,T.,
REFERENCE	O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
REFERENCE	Severy,P., Stange-Thomann,N., Stillwell,J., Stojanovic,N., Stone,C.,
REFERENCE	Subramanian,A., Tesfaye,S., Tichovelsky,N., Torrella-Miller,I.,
REFERENCE	Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
REFERENCE	Ye,W.J., Zhao,J. and Zody,M.
TITLE	Direct Submission
JOURNAL	Submitted (19-NOV-1998) Whitehead Institute/MIT Center for Genome
JOURNAL	Research, 320 Charles street, Cambridge, MA 02141, USA
JOURNAL	3 (bases 1 to 173126)
REFERENCE	Bliren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
REFERENCE	Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
REFERENCE	Castle,A., Cerny,J., Colangelo,M., Collins,S., Collamore,A.,
REFERENCE	Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,
REFERENCE	Donelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C.,
REFERENCE	Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
REFERENCE	Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
REFERENCE	Karats,A., Lehoczy,J., Liu,C., Locke,K., Macdonald,P.,
REFERENCE	Marguis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
REFERENCE	Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychalacky,J.,
REFERENCE	Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,

TITLE Peterson, K., Pollard, V., Riley, R., Roberts, D., Roy, A., Severy, P., Stange, T., Thoman, N., Stojanovic, N., Stone, C., Subramanian, A., Tesfaye, S., Torunella-Miller, I., Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, X., Wyman, D., Ye, R.-J. and Zody, M.

REFERENCE Submitted (14-JAN-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.

TITLE Direct Submission
JOURNAL Submitted (15-JAN-1999) Whitehead Institute/MIT Center for Genome Research, 330 Charles Street, Cambridge, MA 02141, USA
COMMENT On Jan 14, 1999 this sequence version replaced gi:1453855. All repeats were identified using RepeatMasker: Smtl, A.F.A. & Green, P. (1996-1997) <http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Only the first 173126 base pairs of this clone are being submitted
The remainder overlaps accession number AC005988 (WICGR project
L405).

FEATURES	
source	location/Qualifiers
	1. 173126
repeat_region	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="hRPK.299_G-24" /clone_lib="RPCT-11 human" /map="17" /chromosome="17" 464..760
repeat_region	/rpt_family="AluJb" 904..1065
repeat_region	/rpt_family="AluSg/x" complement(1211..1279)
repeat_region	/rpt_family="MIR" 6830..6858
repeat_region	/rpt_family="(TTTTTG)n" 7229..7251
repeat_region	/rpt_family="(TTTA)n" complement(7252..7533)
repeat_region	/rpt_family="AluSg" 9423..9450
repeat_region	/rpt_family="(T)n" complement(9523..9632)
repeat_region	/rpt_family="L2" 9946..10122
repeat_region	/rpt_family="AluJb" 11028..11239
repeat_region	/rpt_family="MBR20" 11471..11519
repeat_region	/rpt_family="(TC)n" 11519..11541
repeat_region	/rpt_family="(TG)n" 11553..11678
repeat_region	/rpt_family="L2" 12346..12447
repeat_region	/rpt_family="(CA)n" complement(12457..12613)
repeat_region	/rpt_family="MIR" complement(12839..12954)

[illegible]

misc_feature /note="assembly_fragment"
115072..128980
/note="assembly_fragment"
misc_feature 129081..142071
/note="assembly_fragment"
misc_feature 142172..165935
/note="assembly_fragment"
misc_feature 166036..190737
/note="assembly_fragment"
misc_feature 190838..214574
/note="assembly_fragment"
clone_end:77
vector_side:right"
BASE COUNT 56656 a 49877 c 51317 g 54418 t 2306 others
ORIGIN
Query Match 2.58; Score 49.2; DB 65; Length 214574;
Best Local Similarity 50.48; Pred. No. 0.18;
Matches 120; Conservative 0; Mismatches 118; Indels 0; Gaps 0;
QY 147 GGACTGCTGTCAGTGCAGCAGCAGAGAGGTCCTTGAAGTGCAGACCCAGTACGAGTGC 206
Db 189074 GGGCAGGGGCGAGAGGCGAGAGGCGAGAGGCGGGGCGAGAGGCGAGAGGCG 189015
QY 207 GGTGCGCCCGCCCTCCCGCCGCTAAACCTGACAGACGACGACAGTACAGTGAAG 266
Db 189014 AGGGGCGAGAGCAGAGGCGAGAGGCGAGAGGCGAGAGGCGAGAGGCGAGAGG 188955
QY 267 GCGGGCCGAGAGGCGCTGCGGAGAGCCCGCGCTACATTGTGTCGACGCGCGCGCGCT 326
Db 188954 AGGGGCGAGAGGCGAGAGGCGAGAGGCGAGAGGCGAGAGGCGAGAGGCGAG 188895
QY 327 GCTGATCCCGAGAGAGGCGCGGTGCTCCAGTCTACTCCGGGAAGTACAGAGCAGC 384
Db 188894 GCAGAAAGCGAGAGAGGCGAGAGGCGAGAGGCGAGAGGCGAGAGGCGAGAGGCG 188837

RESULT 12
LOCUS SCPSCL 11696 bp DNA BCT 30-JUN-1993
DEFINITION S.clavuligerus linear plasmid pSCL (complete sequence).
ACCESSION X54107 S15598
VERSION X54107.1 GI:48758
KEYWORDS Inverted repeat; linear plasmid.
SOURCE Streptomyces clavuligerus.
ORGANISM Plasmid Streptomyces clavuligerus
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 11696)
AUTHORS Roy, K.L.
TITLE Direct Submission
JOURNAL Submitted (26-JUN-1990) Roy K.L., Dept of Microbiology, University
of Alberta, Biological Sciences Bldg Rm M330, Edmonton Alberta,
Canada T6G 2E9
2 (bases 1 to 11696)
Wu, X. and Roy, K.L.
TITLE Complete nucleotide sequence of a linear plasmid from Streptomyces
clavuligerus and characterization of its RNA transcripts
JOURNAL J. Bacteriol. 175 (1), 37-52 (1993)
FEATURES
MEDLINE 93106972
source location/Qualifiers
1. 11696
/organism="Streptomyces clavuligerus"
/plasmid="pSCL"
/strain="NRRL 3585"
/db_xref="taxon:1901"
complement(1446..2237)
/note="ORF1"
/codon_start=1
/transl_table=1
/protein_id="CAA38040.1"
/db_xref="GI:48759"

db_xref="SPTREMBL:005073"
/translation="MAMTRDALLNKGAGVAKTTTTLHIGTLLAAGRRTLLVLDGOG
NLTALKLPRINPAGELTLARAVLEGCGRAGQSLVRRHSENLAVTPSALDFTLPRQ
LHSASKEERSLWVLEHEDGDHCLNDPRLDVPDREILRMADHLLIPVDVEFSI
EALLELGOVOTLVETREAPPRYRGVLINRRVAFPSFGHQAIVTALRLPLPVYGEI
PLRTVAEAKNGOTLADYARTADVARMFRDLAVAGLPEAPAA"
complement(4062..5417)
/note="ORF2"
/codon_start=1
/transl_table=1
/protein_id="CAA38041.1"
/db_xref="GI:581632"
/db_xref="SPTREMBL:005074"
/translation="MSMESEELRAAEVREADEAVLGTLTLPREARDLVLDVSAALPTD
FREPRYGEMLDLASIMARGTCDPDTVTSAYLRKRGDLERIGGAWLHALHDVPTTA
HALRYAEITYAEESRMRALLIAGVRTVTAOREGTAAGAEATAVAGVAREVRGTLAA
DAPPADLIDFLQADDEPRWVTPGLARNDRLMTITAGCGCKSLRLQILVRAAGLH
PWRRARIEPRVLLVDAENASQVPRWLRMAAASAGAEVARGRMHTIYRPGGLD
TAAADRAWLARTVERVRDPLMSIGPLYKLTITSGASDEBARPLMSALMVQASDGA
MLLEAHSPPHAAGVRRRLRPIGSSLWLRMPFGGLSPATGMGAELRLMDWLPWG
PRSEREMPAQFCEGCTWPTAVQYAGNVP IPAGTLTLEQAVAAGLIPQPEWDDQI
G"
complement(6349..6786)
/note="ORF3"
/codon_start=1
/transl_table=1
/protein_id="CAA38042.1"
/db_xref="GI:48761"
/db_xref="SPTREMBL:005076"
/translation="MEALHLRFAQLRQHGEMFDGSDNPVLKYEATITAGLPELPAQ
QPSHNVIPRPSVSKDARCGHELSPHGATGCTVAGDEWLDSCERFRAAEADWH
VGPAPCRHCRHSAODHLIEGPCSKGACVHFASGVGAR"
/transl_table=1
/protein_id="CAA38043.1"
/db_xref="GI:48762"
/db_xref="SPTREMBL:005077"
/translation="MPKTERPLPYMNOVIAHTRDQIDSGELRAGDLISDRIEMG
ISRATQKVTITLAKOGLVEAVOGSTVRAAADLHNSGRRAASVRRGTITTAG
VARIVASIELAPADVAEVLVSQEGAAVIRIRITVYNAENPISASTSWYDGNLAAA
PKLTKTERIKESGMAYLEBOTMTAVYGGDRISARMATEQDQLLIELPAAIKERT
ILRDANSVAVEGVSINGRESIYDVYS"
complement(7921..7951)
/note="promoter for large RNA transcript"
complement(7951..8400)
/note="small RNA transcript"
8114..9367
/note="ORF3"
/codon_start=1
/transl_table=1
/protein_id="CAA38044.1"
/db_xref="GI:581633"
/db_xref="SPTREMBL:005075"
/translation="MLEAPRLSDPTCLARSQSVIAERVAOVAVGVVYAGLAEQ
VEMEREDDAGPTGDPGEGVRRMSERQRELCARAAVAVPGRRVTPSTTADRDLR
TAEVPLTVLVARVAALVEHGDHLVTVARLRDAEWATWAEADVAQDVLVYAR
GRVPELLEVPDLPMPRLLAAAKMSLMDTRRRREWLVRVEDRSADDEVLEALAG
GDSITVCAVDLEMEPEPEPGGWAACVARIALPPROREVELRCTGMMTPATAAR
LGISROSVDRALRFAVTALGSPGTVROBSRAGOPLPAGMGVDRDLRPNTRQDVR
LRAGCARSGEELGEOGLIHNGVAHELHYTRALRSIREMYDQRLDVPAPARSKRPHP
APAGCARSCVGCYLTRTARCAAS"
complement(8411..8442)
/note="promoter for small RNA transcript"
9696..10679
/note="ORF4"
/codon_start=1
/transl_table=1
/protein_id="CAA38045.1"
/db_xref="GI:48764"

```

/db_xref="SPTREMBL:005083"
/translation="MGRRETEPCVPRPQWVLRITLIGAVAVAGVAMAASVTLAELGHA
VGVWMDGRSLNSLPVAVDLALVAGVAVASRLSPMTARSRLAGITLIGAVAGVAVLNA
IGHLVSGSGKVTPLRIAVASVAPPLVAVLGVVLAIPVAGADPGCOOAVLEED
REARITISAHDPADGTADRHDTTARNAPAPPRAPADAVYRPAAPAAARTRAA
DDPAPVPVPAAGTGPRTVATLGDPRDSDQIWERPAPVAPVADRDPLDPCAGTPEP
LLTPDQSRARITMALONTMOTKRIGFEFAGRSASVYNTKKAIRINDGILI"
BASE COUNT      1590 a 4235 c 4179 g 1692 t
ORIGIN
Query Match      2.5%; Score 48.2; DB 3; Length 11696;
Best Local Similarity 55.8%; Pred. No. 0.41;
Matches 92; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

OY 196 GTACAGTGCCTGCTGCGCCCTCCCGCGCGCTAAACCTGACGAGACGAGACAGT 255
    || || || || || || || || || || || || || || || || || || || ||
DB 8198 GTGGCGCAGGTGGGTGTCGGCCGCGTGTGCTGGCTGGCTGAGCAGGTGAGAGCTGG 8257
    || || || || || || || || || || || || || || || || || || || ||
OY 256 GACAGTGAAGGGCGCGGCGAGGGCGCTCGGAGACCCCGCTACATTGTGCGAGCGCGG 315
    || || || || || || || || || || || || || || || || || || || ||
DB 8258 GAGCGGGAGGAGCGACGACGAGCGCGCCGACGAGCGGTGGGTGCTGCTCGGTGG 8317
    || || || || || || || || || || || || || || || || || || || ||
OY 316 CGCGGGCGGCTGCTGATCCCGAGAGCGCGCGTGTCCAGTGG 360
    || || || || || || || || || || || || || || || || || || || ||
DB 8318 TCGGAGCGCTCAGCGCGAGCTGGCGCGCGCGCGGTGCGCGTGG 8362
    || || || || || || || || || || || || || || || || || || || ||

RESULT 13
AC069456      160893 bp      DNA      HTG      25-JUN-2000
LOCUS
DEFINITION
AC069456      Homo sapiens chromosome 4 clone RP11-280K20 map 4, WORKING DRAFT
SEQUENCE
AC069456      22 unordered pieces.
AC069456.2    GI:8705194
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS      Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE        Homo sapiens chromosome 4, clone RP11-280K20
JOURNAL
REFERENCE
AUTHORS      Unpublished
TITLE        2 (bases 1 to 160893)
JOURNAL
REFERENCE
AUTHORS      Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
                Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
                Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
                Campoliano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
                Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
                Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
                Galagan,J., Gardyna,S., Glade,S., Goyette,M., Graham,L.,
                Grand-Pierre,N., Grant,G., Hagos,B., Heathford,A., Horton,L.,
                Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
                Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehoccky,J.,
                Levine,R., Lien,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
                McCathay,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
                Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenka,V., Morrow,J.,
                Murphy,T., Naylor,T., Norman,C.H., O'Connor,T., O'Donnell,P.,
                O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
                Pisanil,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
                Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
                Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
                Teste,S., Theodore,J., Tirelli,A., Travers,M., Trigilio,J.,
                Vassiliev,H., Vieler,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
                Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE        Direct Submission
JOURNAL
COMMENT
Submitted (30-MAY-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 25, 2000 this sequence version replaced gi:8103949.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

```

```

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center: RIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9896
Center clone name: 280_K-20
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye, 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 152447 bases at least Q40
Consensus quality: 156415 bases at least Q30
Consensus quality: 157976 bases at least Q20
Insert size: 163000; agarose-fp
Insert coverage: 158793; sum-of-contigs
Quality coverage: 4.2 in Q20 bases; agarose-fp
Quality coverage: 4.3 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1360: contig of 1360 bp in length
1361 1460: gap of 100 bp
1461 2527: contig of 1067 bp in length
2528 2627: gap of 100 bp
2628 3976: contig of 1349 bp in length
3977 4076: gap of 100 bp
4077 6740: contig of 2664 bp in length
6741 6840: gap of 100 bp
6841 9482: contig of 2642 bp in length
9483 9582: gap of 100 bp
9583 12171: contig of 2589 bp in length
12172 12271: gap of 100 bp
12272 15766: contig of 3495 bp in length.
15767 15866: gap of 100 bp
15867 19727: contig of 3861 bp in length
19728 19827: gap of 100 bp
19828 24165: contig of 4338 bp in length
24166 24265: gap of 100 bp
24266 28290: contig of 4025 bp in length
28291 28390: gap of 100 bp
28391 33226: contig of 4836 bp in length
33227 33326: gap of 100 bp
33327 40113: contig of 6787 bp in length
40114 40213: gap of 100 bp
40214 45498: contig of 5285 bp in length
45499 45598: gap of 100 bp
45599 52958: contig of 7360 bp in length
52959 53058: gap of 100 bp
53059 61457: contig of 8399 bp in length
61458 61557: gap of 100 bp
61558 70503: contig of 8946 bp in length
70504 70603: gap of 100 bp
70604 79766: contig of 9163 bp in length
79767 79866: gap of 100 bp
79867 92993: contig of 13127 bp in length
92994 93093: gap of 100 bp
93094 108057: contig of 14964 bp in length
108058 108157: gap of 100 bp
108158 126261: contig of 18104 bp in length
126262 126361: gap of 100 bp
126362 142680: contig of 16319 bp in length
142681 142780: gap of 100 bp
142781 160893: contig of 18113 bp in length.
Location/Qualifiers
1. 160893
FEATURES
Source

```

				/organism="Homo sapiens"
			/db_xref="taxon:9606"	
			/chromosome="4"	
			/map="4"	
			/clone="RP11-280K20"	
			/clone_lib="RPC1-11 Human Male BAC"	
			1. .1360	
misc_feature		/note="assembly-fragment"	1461. .2527	
misc_feature		/note="assembly-fragment"	2628. .3976	
misc_feature		/note="assembly-fragment"	clone_end:r7	
		vector_side:right"		
misc_feature			4077. .6740	
misc_feature		/note="assembly-fragment"	6841. .9482	
misc_feature		/note="assembly-fragment"	9583. .12171	
misc_feature		/note="assembly-fragment"	12272. .15766	
misc_feature		/note="assembly-fragment"	15867. .19727	
misc_feature		/note="assembly-fragment"	19828. .24165	
misc_feature		/note="assembly-fragment"	24266. .28290	
misc_feature		/note="assembly-fragment"	28391. .33326	
misc_feature		/note="assembly-fragment"	33327. .40113	
misc_feature		/note="assembly-fragment"	40214. .45498	
misc_feature		/note="assembly-fragment"	45599. .52958	
misc_feature		/note="assembly-fragment"	53059. .61457	
misc_feature		/note="assembly-fragment"	61558. .70503	
misc_feature		/note="assembly-fragment"	70604. .79766	
misc_feature		/note="assembly-fragment"	79867. .92993	
misc_feature		/note="assembly-fragment"	93094. .108057	
misc_feature		/note="assembly-fragment"	108158. .126261	
misc_feature		/note="assembly-fragment"	126362. .142680	
		/note="assembly-fragment"		
		clone_end:sp6		
		vector_side:left"		
misc_feature		/note="assembly-fragment"	142781. .160893	
BASE COUNT	51136 a	28485 c	27792 g	51377 t
ORIGIN			2103	others
Query Match	2.4%	Score 47.4;	DB 74;	Length 160893;
Best Local Similarity	46.3%	Pred. No. 0.5;		
Matches 156;	Conservative	0;	Mismatches 181;	Indels 0;
			Gaps	0.0;
Db	103	GGCCGGGAGTCTCCGGCTAGCTGCATTCACAGACCGTCTGTGTGACCTTGGTCACTTGGTCACTGAC	162	
Db	91957	GGCTGCAGTCTCGCCCTTGGCGCGGACCTCTACGCCCTTGGGTGGTGTGATGGCATTG	92016	
Db	163	AGCGACGAAGAGGTCTTGGAGTTCGAGACCCAGTATGAGAGTCCGGTCCGCCGCTCCCC	222	
Db	92017	AGTGCCTGTGGACACAGGGGCGCGCTCGTGTGGGGGCGGCTGTGGGGCGGCGCTCTT	92076	
Db	223	GGCGCGGCTAAACCTGAGACGACGACGACGACGACGACGACGACGACGACGACGACGACG	282	
Db	92077	TGGGAGGCTCCCGCGCACAGATCCCAACGAGGAGGGGTGAGAGCTCAGGCAATGGCGGCT	92136	

Oy	283	GCGGGACCCCCCGGTCAATTGTCGACAGCCGCCGCCGCTGCTGANTCCCGAGAG	342
Db	92137	GCAAGTTCGGAGGCCCTGCCTCGCGGAGCAGCTAAAGCCCCTGTGAATAATCACGGCA	92196
Oy	343	GCAGCCGCTGGTCCCATGTACTCCGGGAGCTACAGACACCCTCAACCTATTCCAGAT	402
Db	92197	GCGCTGTGGGCGGCACTGCTGGGGGACCAGCAACACCTCCGACCGCTGGCCCGG	92256
Oy	403	AATTCATCCCTTTGAAACGTGCCCTTTACAGAGCCTG	439
Db	92257	TGCTAAGCCCCCTATTGCTCGGGGCCCTTTTAGGGCCGG	92293

RESULT	14
AC022573/c	
LOCUS	AC022573 77836 bp DNA
DEFINITION	Homo sapiens clone Rpl1-12N11, LOW-PASS SEQUENCE SAMPLING.
ACCESSION	AC022573
VERSION	AC022573.2 GI:7144847
KEYWORDS	HTG; HTGS_PHASE0.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Eukaryota: Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 77836)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-12N11
Unpublished
2 (bases 1 to 77836)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson, S., Baldwin, J., Barna, N., Beckert, R., Beda, F., Boguslavsky, L., Boukhgalter, B., Brown, A., Burnett, G., Castle, A., Chepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Deardaroff, K., Dewar, K., Domino, M., Doyle, M., Fenster, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Gargagan, J., Gardyna, S., Grant, G., Hagos, B., Heatford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lenders, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, R., Macdonald, P., Marquis, N., McEwan, J., Mccurk, A., McKernan, K., Mccheeters, R., Meldrum, J., Meunus, L., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Olivari, T. M., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tirrell, A., Vassiliiev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

TITLE Direct Submission
JOURNAL Whitehead Institute/MIT Center for Genome
Submitted (06-FEB-2000)
RESEARCH 320 Charles Street, Cambridge, MA 02141, USA
On Mar 3, 2000 this sequence version replaced g1:6510780.
COMMENT

ALL repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L3299

Center clone name: 12_N_11

NOTE: This record contains 89 individual

* sequencing reads that have not been assembled into contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allow
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will

```
* be preserved.
* 1 762: contig of 762 bp in length
* 763 862: gap of 100 bp
* 863 1633: contig of 771 bp in length
* 1634 1733: gap of 100 bp
* 1734 2505: contig of 772 bp in length
* 2506 2605: gap of 100 bp
* 2606 3373: contig of 768 bp in length
* 3374 3473: gap of 100 bp
* 3474 4273: contig of 800 bp in length
* 4274 4373: gap of 100 bp
* 4374 5153: contig of 780 bp in length
* 5154 5253: gap of 100 bp
* 5254 6016: contig of 763 bp in length
* 6017 6116: gap of 100 bp
* 6117 6881: contig of 765 bp in length
* 6882 6981: gap of 100 bp
* 6982 7736: contig of 757 bp in length
* 7737 7838: gap of 100 bp
* 7839 8600: contig of 762 bp in length
* 8601 8700: gap of 100 bp
* 8701 9464: contig of 764 bp in length
* 9465 9564: gap of 100 bp
* 9565 10338: contig of 775 bp in length
* 10340 10439: gap of 100 bp
* 10440 11223: contig of 784 bp in length
* 11224 11323: gap of 100 bp
* 11324 12111: contig of 788 bp in length
* 12112 12211: gap of 100 bp
* 12212 12970: contig of 759 bp in length
* 12971 13070: gap of 100 bp
* 13071 13870: contig of 800 bp in length
* 13871 13970: gap of 100 bp
* 13971 14724: contig of 754 bp in length
* 14725 14824: gap of 100 bp
* 14825 15596: contig of 772 bp in length
* 15597 15696: gap of 100 bp
* 15697 16456: contig of 760 bp in length
* 16457 16556: gap of 100 bp
* 16557 17314: contig of 758 bp in length
* 17315 17414: gap of 100 bp
* 17415 18176: contig of 762 bp in length
* 18177 18276: gap of 100 bp
* 18277 19036: contig of 760 bp in length
* 19037 19136: gap of 100 bp
* 19137 19931: contig of 795 bp in length
* 19932 20031: gap of 100 bp
* 20032 20823: contig of 792 bp in length
* 20824 20923: gap of 100 bp
* 20924 21708: contig of 785 bp in length
* 21709 21808: gap of 100 bp
* 21809 22582: contig of 774 bp in length
* 22583 22682: gap of 100 bp
* 22683 23443: contig of 761 bp in length
* 23444 23543: gap of 100 bp
* 23544 24326: contig of 783 bp in length
* 24327 24426: gap of 100 bp
* 24427 25200: contig of 774 bp in length
* 25201 25300: gap of 100 bp
* 25301 26065: contig of 765 bp in length
* 26066 26165: gap of 100 bp
* 26166 26922: contig of 757 bp in length
* 26923 27022: gap of 100 bp
* 27023 27796: contig of 774 bp in length
* 27797 27896: gap of 100 bp
* 27897 28680: contig of 784 bp in length
* 28681 28780: gap of 100 bp
* 28781 29571: contig of 791 bp in length
* 29572 29671: gap of 100 bp
* 29672 30473: contig of 802 bp in length
* 30474 30573: gap of 100 bp
* 30574 31338: contig of 765 bp in length
* 31339 31438: gap of 100 bp

31439 32217: contig of 779 bp in length
32218 32317: gap of 100 bp
32318 33093: contig of 776 bp in length
33094 33193: gap of 100 bp
33194 33956: contig of 763 bp in length
33957 34056: gap of 100 bp
34057 34818: contig of 762 bp in length
34819 34918: gap of 100 bp
34919 35811: contig of 893 bp in length
35812 35911: gap of 100 bp
35912 36663: contig of 752 bp in length
36664 36763: gap of 100 bp
36764 37553: contig of 790 bp in length
37554 37653: gap of 100 bp
37654 38412: contig of 759 bp in length
38413 38512: gap of 100 bp
38513 39258: contig of 747 bp in length
39260 39359: gap of 100 bp
39360 40144: contig of 785 bp in length
40145 40244: gap of 100 bp
40245 41027: contig of 783 bp in length
41028 41127: gap of 100 bp
41128 41907: contig of 780 bp in length
41908 42007: gap of 100 bp
42008 42777: contig of 770 bp in length
42778 42877: gap of 100 bp
42878 43682: contig of 805 bp in length
43683 43782: gap of 100 bp
43783 44559: contig of 777 bp in length
44560 44659: gap of 100 bp
44660 45429: contig of 770 bp in length
45430 45529: gap of 100 bp
45530 46280: contig of 751 bp in length
46281 46380: gap of 100 bp
46381 47146: contig of 766 bp in length
47147 47246: gap of 100 bp
47247 48008: contig of 762 bp in length
48009 48108: gap of 100 bp
48109 48881: contig of 773 bp in length
48882 48981: gap of 100 bp
48982 49762: contig of 781 bp in length
49763 49862: gap of 100 bp
49863 50654: contig of 792 bp in length
50655 50754: gap of 100 bp
50755 51535: contig of 781 bp in length
51536 51635: gap of 100 bp
51636 52451: contig of 816 bp in length
52452 52551: gap of 100 bp
52552 53307: contig of 756 bp in length
53308 53407: gap of 100 bp
53408 54173: contig of 766 bp in length
54174 54273: gap of 100 bp
54274 55037: contig of 764 bp in length
55038 55137: gap of 100 bp
55138 55904: contig of 767 bp in length
55905 56004: gap of 100 bp
56005 56753: contig of 749 bp in length
56754 56853: gap of 100 bp
56854 57637: contig of 784 bp in length
57638 57737: gap of 100 bp
57738 58519: contig of 782 bp in length
58520 58619: gap of 100 bp
58620 59414: contig of 795 bp in length
59415 59514: gap of 100 bp
59515 60301: contig of 787 bp in length
60302 60401: gap of 100 bp
60402 61193: contig of 792 bp in length
61194 61293: gap of 100 bp
61294 62068: contig of 775 bp in length
62069 62168: gap of 100 bp

Query Match 2.48; Score 47.2; DB 67; Length 77836;
Best Local Similarity 43.0%; Pred. No. 0.6;
```

	Matches	148:	Conservative	0:	Mismatches	196:	Indels	0:	Gaps	0:
QY	30	GGGACGTGTGTCGAGATCCCGCGGTGGCCCGAGCGCTGGAGAGCCCGAGCGCCCTGTG	89							
Db	35624	GGGCGGGGGGGCGCGNNNNNGGGGGGGCGCGCGCGCCGCGCGCGCGGGGGGGGG	35655							
QY	90	CCGGTGTCTCTCGCGCCCGGCACTCTCCGGGCTTGGGCTCATTTCCAGACACCGTCTTGTGA	149							
Db	35564	CCGGCGCCCGCGCGGG	35505							
QY	150	CTTGTGCTAGTCAGCAGCAGAGAGGCTTCTGGAAGTCGCGAGACCCAGTACGATGCGCGGT	209							
Db	35504	GGGGGGGGGGGGCGCGCGNNGGNCNGGGGGGGGGCGCGGGGGGGGGGGGGGGGGGG	35445							
QY	210	CGCCCGCTCCCGCGCGCGCTTAACTTGAGCAGGACACGCACTGACATGTAAGAGGGCG	269							
Db	35444	CCGCGGGGGCGCGGGGNGGGGGCGCGGCGNNGGNGGNGGGGGGGGGGGGGGGGGGGGG	35385							
QY	270	GAGCGAGGGGCTCGCGAGAGCCCGCGTACATTGGTGCAGCGCGCGCGGCGGCTGCT	329							
Db	35384	GGGGGGGGGGGNNNGGGGGGGGGGGCGCGCGGNNNGGGGGGGGGGGGCGGGGGNGGG	35325							
QY	330	GGATCCGGAGAGGGCGCGGTGTCTCCCACTACTCCGGGAAGG	373							
Db	35324	GGGCGCGGNCGGGGGGGGGGCGNNNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	35281							

Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
 Thomas, S., Usmani, K., Vasquez, L., Veta, V., Villalón, D., Vinson, R.,
 Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
 Wellington, S., Williams, G., Williamson, A., Wlezyk, R., Wooden, S.,
 Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.
 and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 147225)
 Morley, K.C.
 Direct Submission
 Submitted (05-FEB-2000) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Nov 3, 2000 this sequence version replaced g1:9438272.
 ----- Genome Center -----
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information -----
 Center project name: HADJ
 Center clone name: RP11-430L6
 ----- Summary Statistics -----
 Sequencing vector: M13, L08821

RESULT	15
LOCUS	AC022498
DEFINITION	AC022498 147225 bp DNA HTG 04-NOV-2000
ACCESSION	AC022498
VERSION	AC022498.12 GI:11079244
KEYWORDS	HTG; HTGS-PHASE1; HTGS-DRAFT.
SOURCE	human.
ORGANISM	Homio sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS	1 (bases 1 to 147225)
	Muzny,D.M., Adams,C., Adio-Oduola,B., All-oman,F.R., Allen,C., Alibrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbara,J., Benton,J., Blmage,K., Blankenburg,K., Bonnh,D., Bouk,J., Buche,S., Bivleva,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C., Burich,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Chatter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleeland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Day,Carrill,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dlnh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotlo,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevar,W., Gnnathne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Hayla,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Huyl,J.S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korah,J., Koyar,C., Krivoyic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Licharge,O., Iieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Lund,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Matlnez,E., Massey,E., Mathney,E., McLeod,M.P., Meador,M., Metizer,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwkenko,S., Ough,M., Okwomu,G., Ogunyeye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Prlms,E., Pul,L.L., Qulles,M., Ren,Y., Rives,M., Rojs,A., Rojudoan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostlati,N., Sisson,I., Sodergren,E., Sonake,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,

Tang, H., Tsey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Washington, S., Williams, G., Williamson, A., Wlarczyk, R., Woodén, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 147225)
Worley, K.C.

Direct Submission
Submitted (05-FEB-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 3, 2000 this sequence version replaced gi:9438272.

Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

Project Information
Center project name: HAD1-430L16
Center clone name: RP11-430L16

Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-primer; BodyPip; Big Dye; 90% of reads
Assembly program: Phrap; Variation 0.990329
Consensus quality: 119390 bases at least Q40
Consensus quality: 134432 bases at least Q30
Consensus quality: 138749 bases at least Q20
Estimated insert size: 138179; sum-of-configs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 6x in Q20 bases; sum-of-configs estimation

* NOTE: Estimated insert size may differ from sequence length.
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_drift_data.html)
* NOTE: This is a "working draft" sequence. It currently
* consists of 29 configs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the configs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1	10678: contig of 10678 bp in length
*	10679: gap of unknown length
*	10779: contig of 17118 bp in length
*	27897: gap of unknown length
*	27997: contig of 12120 bp in length
*	40117: gap of unknown length
*	40217: contig of 13626 bp in length
*	53843: gap of unknown length
*	53943: contig of 6191 bp in length
*	60134: gap of unknown length
*	60234: contig of 6720 bp in length
*	66953: gap of unknown length
*	66954: gap of unknown length
*	67054: gap of unknown length
*	72921: contig of 3868 bp in length
*	73022: gap of unknown length
*	81158: contig of 8137 bp in length
*	81159: gap of unknown length
*	81258: gap of unknown length
*	88668: contig of 7410 bp in length
*	88669: gap of unknown length
*	88769: contig of 5954 bp in length
*	94722: gap of unknown length
*	94823: contig of 5462 bp in length
*	100285: gap of unknown length
*	100385: contig of 4995 bp in length
*	105380: gap of unknown length
*	105480: contig of 4825 bp in length
*	110505: gap of unknown length
*	110405: contig of 6402 bp in length
*	116807: gap of unknown length


```
* 116907 120759: contig of 3853 bp in length
* 120760 120859: gap of unknown length
* 120860 125348: contig of 4489 bp in length
* 125349 125448: gap of unknown length
* 125449 127226: contig of 1778 bp in length
* 127227 127326: gap of unknown length
* 127327 129574: contig of 2248 bp in length
* 129575 129674: gap of unknown length
* 129675 131343: contig of 1669 bp in length
* 131344 131443: gap of unknown length
* 131444 133042: contig of 1599 bp in length
* 133043 133142: gap of unknown length
* 133143 135153: contig of 2011 bp in length
* 135154 135253: gap of unknown length
* 135254 137006: contig of 1753 bp in length
* 137007 137106: gap of unknown length
* 137107 138550: contig of 1444 bp in length
* 138551 138651: gap of unknown length
* 138652 140149: contig of 1499 bp in length
* 140150 140249: gap of unknown length
* 140250 141631: contig of 1382 bp in length
* 141632 141731: gap of unknown length
* 141732 143173: contig of 1442 bp in length
* 143174 143273: gap of unknown length
* 143274 144588: contig of 1315 bp in length
* 144589 144688: gap of unknown length
* 144689 145782: contig of 1094 bp in length
* 145783 145882: gap of unknown length
* 145883 147225: contig of 1343 bp in length.
```

FEATURES

Source

```
1. 147225
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="3"
   /clone="RP11-430L16"
```

BASE COUNT 41313 a 31018 c 30751 g 41299 t 2844 others

ORIGIN

Query Match 2.4%; Score 47; DB 67; Length 147225;

Best Local Similarity 49.3%; Pred. No. 0.63; Mismatches 151; Indels 1; Gaps 1;

Matches 148; Conservative 0; Mismatches 151; Indels 1; Gaps 1;

```
OY 25 CTGAGGGGACCTGTCCGAGGTCCGCGGTGCGAGCGCTCCGAGAGCCCGGAGCGCC 84
Db 88345 CGGCGGGGGCGGGGGTGGGTGGCGGTGCGCGGGCGCGGGCGCGCGCGCG 88404

OY 85 CGTGGCCGCTCTCTCCGCGCCGCGACGTCCTGAGCTCATTCAGACACCGTGCTT 144
Db 88405 CGGCGGGGGCGGGGGCGGCTTTCGGGGCGCGCGGGCGCGCGGGGGCGCGGT 88464

OY 145 GTGACTTGTGCTGACGTGACGAGAGAGTCTTGGAGTCCGAGACCCAGTAGAGTG 204
Db 88465 GCCCGGTGGGCGCGGGCGGGGGCGGGGTGGTGGGGGGCGGTCCGCGCGCGCG 88524

OY 205 CCGGTCGCCCGCTCCCGCGCGCGCTAAACTGACGACGACAGACAGTGAAGTAA 264
Db 88525 GCGGGGGGGCGG -CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTG 88583

OY 265 GGGCGGGCGCGAGGGGCTTGGGGAGCCCGCGTACATTGGTGCGAGCGGGCGGGCGG 324
Db 88584 GCGGTGGCGGCGGGCGGGGTGGGTGCGCGCGCGCGCGGTGCGGGCGCGGTGCGG 88643
```

Search completed: September 11, 2001, 12:44:19
Job time: 3951 sec

This Page Blank (except for)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 11, 2001, 11:41:48 ; Search time 1471.75 Seconds
(without alignments)
12498.898 Million cell updates/sec

Title: US-09-192-611-1

Perfect score: 1946
Sequence: 1 ACAGTGTGGAGATGCGGA.....CGCGATTCGCGCCGACTCGAG 1946

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est25.*
26: gb_est26.*
27: gb_est27.*
28: gb_est28.*
29: gb_est29.*
30: gb_est30.*
31: gb_est31.*
32: gb_est32.*
33: gb_est33.*
34: gb_est34.*
35: gb_est35.*
36: gb_est36.*
37: gb_est37.*
38: gb_est38.*
39: gb_est39.*
40: gb_est40.*
41: gb_est41.*
42: gb_est42.*
43: gb_est43.*
44: gb_est44.*
45: gb_est45.*
46: gb_est46.*
47: gb_est47.*
48: gb_est48.*
49: gb_est49.*
50: gb_est50.*
51: gb_est51.*
52: gb_est52.*
53: gb_est53.*
54: gb_est54.*
55: gb_est55.*
56: gb_est56.*
57: gb_est57.*
58: gb_est58.*
59: gb_est59.*
60: gb_est60.*
61: gb_est61.*
62: gb_est62.*
63: gb_est63.*
64: gb_est64.*
65: gb_est65.*
66: gb_est66.*
67: gb_est67.*
68: gb_est68.*
69: gb_est69.*
70: gb_est70.*
71: gb_est71.*
72: gb_est72.*
73: gb_est73.*
74: gb_est74.*
75: gb_est75.*
76: gb_est76.*
77: gb_est77.*
78: gb_est78.*
79: gb_est79.*
80: gb_est80.*
81: gb_est81.*
82: gb_est82.*
83: gb_est83.*
84: gb_est84.*
85: gb_est85.*
86: gb_est86.*
87: gb_est87.*
88: gb_est88.*
89: gb_est89.*
90: gb_est90.*
91: gb_est91.*
92: gb_est92.*
93: gb_est93.*
94: gb_est94.*
95: gb_est95.*
96: gb_est96.*
97: gb_est97.*
98: gb_est98.*
99: gb_est99.*
100: gb_est100.*
101: gb_est101.*
102: gb_est102.*
103: gb_est103.*
104: gb_est104.*
105: gb_est105.*
106: gb_est106.*
107: gb_est107.*
108: gb_est108.*
109: gb_est109.*
110: gb_est110.*
111: gb_est111.*
112: gb_est112.*
113: gb_est113.*
114: gb_est114.*
115: gb_est115.*
116: gb_est116.*

44: em_esthum10.*
45: em_esthum11.*
46: em_esthum12.*
47: em_esthum13.*
48: em_esthum14.*
49: em_esthum15.*
50: em_esthum16.*
51: em_esthum17.*
52: em_esthum18.*
53: em_esthum19.*
54: em_esthum20.*
55: em_esthum21.*
56: em_esthum22.*
57: em_esthum23.*
58: em_esthum24.*
59: em_esthum25.*
60: em_esthum26.*
61: em_esthum27.*
62: em_esthum28.*
63: em_esthum29.*
64: em_esthum30.*
65: em_esthum31.*
66: em_esthum32.*
67: em_esthum33.*
68: em_esthum34.*
69: em_esthum35.*
70: em_esthum36.*
71: em_esthum37.*
72: em_esthum38.*
73: em_esthum39.*
74: em_esthum40.*
75: em_esthum41.*
76: em_esthum42.*
77: em_esthum43.*
78: em_esthum44.*
79: em_esthum45.*
80: em_esthum46.*
81: em_esthum47.*
82: em_esthum48.*
83: em_esthum49.*
84: em_esthum50.*
85: em_esthum51.*
86: em_esthum52.*
87: em_esthum53.*
88: em_esthum54.*
89: em_esthum55.*
90: em_esthum56.*
91: em_esthum57.*
92: em_esthum58.*
93: em_esthum59.*
94: em_esthum60.*
95: em_esthum61.*
96: em_esthum62.*
97: em_esthum63.*
98: em_esthum64.*
99: em_esthum65.*
100: em_esthum66.*
101: em_esthum67.*
102: em_esthum68.*
103: em_esthum69.*
104: em_esthum70.*
105: em_esthum71.*
106: em_esthum72.*
107: em_esthum73.*
108: em_esthum74.*
109: em_esthum75.*
110: em_esthum76.*
111: em_esthum77.*
112: em_esthum78.*
113: em_esthum79.*
114: em_esthum80.*
115: em_esthum81.*
116: em_esthum82.*

```
117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*
```

```
190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1333.6	68.5	1354	192	AK005947	AK005947 Mus muscu
2	443.2	22.8	871	175	BG291400	BG291400 602387277
3	418.6	21.5	962	165	BE260911	BE260911 601153835
4	413.6	21.3	946	139	BE741369	BE741369 601594227
5	392.4	20.2	544	165	BE253049	BE253049 601116953
6	336.2	17.3	636	248	AZ720215	AZ720215 RPCT-24-1
7	325.8	16.7	904	141	BE868967	BE868967 601444727
8	309.6	15.9	799	172	BG026071	BG026071 602292263
9	297.6	15.3	907	141	BE867810	BE867810 601443775
10	264.2	13.6	525	31	AV610897	AV610897 AV610897
11	258.8	13.3	669	147	BF130650	BF130650 601894605
12	249.8	12.8	352	6	AA351060	AA351060 EST58647
13	224.6	11.5	239	123	BB018959	BB018959 BB018959
14	218.2	11.2	616	153	BG392671	BG392671 602410920
15	206.6	10.6	544	144	BF076210	BF076210 225720 MA
16	193	9.9	245	27	AV270183	AV270183 AV270183
17	189.6	9.7	215	109	AV040695	AV040695 AV040695
18	187	9.6	474	104	A1954626	A1954626 WQ34C08.x
19	183.8	9.4	457	174	BG150572	BG150572 7K02H09.x
20	177.2	9.1	608	175	BG289958	BG289958 602381375
21	157	8.1	673	175	BG283797	BG283797 602407512
22	152.6	7.8	675	154	BG483779	BG483779 602503425
23	151.4	7.8	219	1	AA063239	AA063239 ZM02101.S
24	146.2	7.5	291	6	AA352196	AA352196 EST60152
25	146.2	7.5	531	122	AAW66801	AAW66801 EST378875
26	122	6.3	607	116	AAW46292	AAW46292 hc80907.x
27	119.8	6.2	388	119	AAW67539	AAW67539 bB44G06.Y
28	117.6	6.0	667	230	AA058438	AA058438 RPCT-11-4
29	98	5.0	366	5	AA300063	AA300063 EST12698
30	94.8	4.9	237	13	AA919081	AA919081 O178C04.S
31	80	4.1	618	7	AA414516	AA414516 vD09e01.S
32	70.6	3.6	617	138	BE680653	BE680653 dF83C04.Y
33	64.4	3.3	597	115	AAW420637	AAW420637 fJ84D09.x
34	62	3.2	568	138	BE678161	BE678161 dF83C04.x
35	59.6	3.1	925	219	CNS0091P	AA053013 Drosophila
36	56.4	2.9	311	5	AA300064	AA300064 EST12699
37	56.4	2.9	716	155	BC6562954	BC6562954 602581118
38	52.6	2.7	932	219	CNS0072Q	AL066742 Drosophila
39	52.4	2.7	1415	76	AW727479	AW727479 GA_Ea001
40	52	2.7	925	219	CNS0091P	AL053013 Drosophila
41	50.8	2.6	1101	219	CNS0173X	AL108460 Drosophila
42	50.6	2.6	776	219	CNS010RY	AL099352 Drosophila
43	50.6	2.6	908	219	CNS006B4	AL064031 Drosophila
44	50.4	2.6	935	219	CNS006XK	AL066051 Drosophila
45	49.6	2.5	440	23	A1671575	A1671575 WB33H03.x

ALIGNMENTS

RESULT 1
LOCUS AK005947
DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:1700013F11, full insert sequence.
ACCESSION AK005947
VERSION AK005947.1
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone:1700013F11.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (sites)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning

JOURNAL

Methods Enzymol. 303, 19-44 (1999)

REFERENCE

2 (sites)
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

TITLE

Normalization and subraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL

Genome Res. 10 (10), 1617-1630 (2000)

REFERENCE

3 (sites)
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Kikuchi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Okazaki, Y., Muramatsu, M., Inoue, Y. and Hayashizaki, Y.

TITLE

RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer

JOURNAL

Genome Res. 10 (11), 1757-1771 (2000)

REFERENCE

4 (sites)
The RIKEN Genome Exploration Research Group Phase II Team and FANTOM Consortium

TITLE

Functional annotation of a full-length mouse cDNA collection

JOURNAL

Nature 409, 685-690 (2001)

REFERENCE

5 (bases 1 to 1354)
Aachari, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hirose, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Teijima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

TITLE

Direct Submission

JOURNAL

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.go.jp, URL: http://genome-gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT

Please visit our web site (<http://genome-gsc.riken.go.jp/>) for further details.

FEATURES

source
1. 1354
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="MGI:1329015"
/db_xref="MGI:1906296"
/clone="1700013F11"
/sex="male"
/tissue="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
/note="putative"

CDS

```
/codon_start=2
/protein_id="BAB24331.1"
/db_xref="GI:12838793"
/translation="SVSGDGTTEGTMSEVPMRPRSETRGARGCRPARARLIP
bTVLVLDVSDSEVLEVADPVAVPARLPAPAKPEQDSDSDEGAEPAGAPRLV
RRRRRLDLPGEAPVVPVYSGKVQSLNLIIPNSSLJLKPCSPEDDEALJTSGPS
EDDALFSGSPMRKLLRKKEKEKMEEPDQDIPLPSPSSNKRKRTKLEALQKRE
VNRKLDLRSCLSKPOHOSPALOSTDEVVLEVPVLPSPSRLETKIRCADLVLP
VMSSEPIQVVDHMANHGLGSPNRILLIFGSEPISTPATSPRLTKIGVADITICVIVAS
SSEATETSELRILRQGREKHOMLEISLSPDPLKVLMSHYEAMLSGKHLSFFPDG
TMSSEKELPADLGESEGLIEYWG"
BASE COUNT      312 a      368 c      396 g      278 t
ORIGIN

Query Match      68.5%; Score 1333.6; DB 192; Length 1354;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1347; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
```

```
Db 783 TCTTTACACTCAAGATCCGGGTGCCGGCTGACCTAGTAGACATCTGTCAGATGTCG 842
QY 842 AGCCCTTCGAGATGTCGTGATCACATGGCCAAATCATCTTGGGGTGTCCAAACAGGA 901
Db 843 AGCCCTTCGAGATGTCGTGATCACATGGCCAAATCATCTTGGGGTGTCCAAACAGGA 902
QY 902 TTTCTTTGGCTTTTGGAGAGATGAATCTCTCTACTGCGACCCCTAGTACCTTAAGC 961
Db 903 TTTCTTTGGCTTTTGGAGAGATGAATCTCTCTACTGCGACCCCTAGTACCTTAAGC 962
QY 962 TTGAGTGCCTGACATCATTTGATTGTGTGCTGTAGCAAGCTTTGAGAGCCACAGAGA 1021
Db 963 TTGAGTGCCTGACATCATTTGATTGTGTGCTGTAGCAAGCTTTGAGAGCCACAGAGA 1022
QY 1022 CATCCAGAGAGCTCCGGCTCCGGGTGAGGGAGGAAGCAACCAAGATGTGAGATCT 1081
Db 1023 CATCCAGAGAGCTCCGGCTCCGGGTGAGGGAGGAAGCAACCAAGATGTGAGATCT 1082
QY 1082 CACTGTCTCCTGATTCCTCTTAAAGTTTCTCATGTGCACACTATGAGAGAGCCATGGAG 1141
Db 1083 CACTGTCTCCTGATTCCTCTTAAAGTTTCTCATGTGCACACTATGAGAGAGCCATGGAG 1142
QY 1142 TCTGTGACACAAGCTCTCTCTTTCTTTGATGGACAAAGCTTTGACGCAAGAGCTGC 1201
Db 1143 TCTGTGACACAAGCTCTCTCTTTCTTTGATGGACAAAGCTTTGACGCAAGAGCTGC 1202
QY 1202 CAGCTGATCTGGGCTCGGAATTCGGAGATCTCATGCAAGTCTGGGGTGAAGCTCTCAC 1261
Db 1203 CAGCTGATCTGGGCTCGGAATTCGGAGATCTCATGCAAGTCTGGGGTGAAGCTCTCAC 1262
QY 1262 CTGTTCGGACGCAAGGCAAGACATGAGACAATAGCTCCCAATTTATTATTGTGATTT 1321
Db 1263 CTGTTCGGACGCAAGGCAAGACATGAGACAATAGCTCCCAATTTATTATTGTGATTT 1322
QY 1322 TTGCCCCCAATAGGCTTAACAGAACTGATTT 1353
Db 1323 TTGCCCCCAATAGGCTTAACAGAACTGATTT 1354
```

```
RESULT 2
LOCUS BG291400 871 bp mRNA EST 21-FEB-2001
DEFINITION 602387277F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:451614 5',
mRNA sequence.
ACCESSION BG291400
VERSION BG291400.1 GI:13049266
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NIH-MGC http://mgi.ncl.nih.gov/
NATIONAL INSTITUTES OF HEALTH, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cygabs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLML0407 Row: 1 Column: 19
High quality sequence stop: 729.
Location/Qualifiers
1..871
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:451614"
/clone_1lb="NIH_MGC_93"
/tissue_type="transitional cell papilloma, cell line"
```

FEATURES

source

DB	534	TCATTGACGTGTGGTACTACAAAGTTCTCCAGAGGCCACAGAGAGTCCCAACAGCTTC	593
OY	1037	GGCTCCGGGTGTCAGGGGAAAGAAAACACCAGATGTTGGAGATCTTCACTGTCTCTGATT	1096
DB	594	AGCTCCGGGTGTCAGGGAAAGAGAAAGCCACAGACTGGAGATCTCACTGTCTCGAGATT	653
OY	1097	CTCCCTTTAAAGTCTTCATGTCACACTATAGAGAAAGCCATGGGACTCTCTGGACACAAAC	1156
DB	654	CCCCCTTTAAAGAACTTCATGTCCTCCTATAGAGG-AGCCCTGGAGCCGGTGGAGGACGAGC	712
OY	1157	TCTCTCTCTCTT	1169
DB	713	TCTCTCTCTCTT	725
RESULT	4		
LOCUS	BE741369	946 bp	EST
DEFINITION	601594227/1 NIH_MGC_9	Homo sapiens cDNA clone IMAGE:3848270 5'	15-SEP-2000
ACCESSION	BE741369		
VERSION	BE741369.1	GI:10155361	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 946)		
JOURNAL	NIH-MGC http://mhc.nci.nih.gov/		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: csapbs-remail.nih.gov		
	Tissue Procurement: DCTD/DBP		
	cDNA Library Preparation: Ling Hong/Rubin Laboratory		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/ULNL at: image.ln1.gov		
	Plate: LHCMB11 row: e column: 07		
	High quality sequence stop: 751.		
FEATURES	Location/Qualifiers		
Source	1..946		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:3948270"		
	/clone_lib="NIH_MGC_9"		
	/tissue_type="adenocarcinoma cell line"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: ovary; Vector: POTB7; Site_1: XhoI; Site_2:		
	EcoRI; cDNA made by oligo-dT priming. Directionality		
	cloned into EcoRI/XhoI sites using the following 5'		
	adaptor: GGCACAG(G). Size-selected >500bp for average		
	insert size 1.8kb. Library constructed by Ling Hong in		
	the Laboratory of Gerald M. Rubin (University of		
	California, Berkeley) using ZAP-cDNA synthesis kit		
	(Stratagene) and Superscript II RT (Life Technologies)."		
BASE COUNT	183 a	274 c	266 g
ORIGIN			223 t
Query Match	21.3%	Score 413.6;	DB 139; Length 946;
Best Local Similarity	79.4%	Pred. No. 1e-102;	
Matches	508;	Conservative 0;	Mismatches 119; Indels 13; Gaps 1;
OY	725	GCACAGATGATGAGTGTGCTAGTGAAGGCGCTGTCTTGGCACAGAGCTTCGACTCT	784
DB	13	GCCAGAGAGATGAAGTGGTCTTGAGGAAGGCGCCACCCCTCCAGAGAGCCCGACTCT	72
OY	765	TTACACTCAAGTCCGGGTCCGGGCTGACCTTGTGAGACTGCTGTGAGATGTGGAGC	844
DB	73	TCCACCTCAAAATCCGTTGCGCGGTGACACTGTGATGATGTCCTCCAGATGTGGAGC	132

QY	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
QY	845	CCCTTCAGAAATGTGGTGGATACATGGCCATTCATCTTTGGGGTGTCTCCAAACAGATTC	BE253049	544 bp	mRNA	EST	13-JUL-2000					
Db	133	CCCTTCAGAGTGTGGTGGACCAATGGCCACCTTGGGGTGTCTCCAAACAGATTC	60111695.F1	NIH_MGC_16	Homo sapiens cDNA clone	IMAGE:3357421	5'					
QY	905	TTTTGCTTTTGGAGAGAGTGAATCTGTCTTACCTGTGCACCCCTAGTAACTTAAAGCTTG	BE253049	544 bp	mRNA	EST	13-JUL-2000					
Db	193	TTTTGCTTTTGGAGAGAGTGAATCTGTCTTACCTGTGCACCCCTAGTAACTTAAAGCTTG	60111695.F1	NIH_MGC_16	Homo sapiens cDNA clone	IMAGE:3357421	5'					
QY	965	GAGTGGCTGACATCATTTGATTTGTGTGTGTGCTAGCAAGCTCTTCAGAGGCCACAGACAT	BE253049	544 bp	mRNA	EST	13-JUL-2000					
Db	253	GAGTGGCTGACATCATTTGATTTGTGTGTGTGCTAGCAAGCTCTTCAGAGGCCACAGACAT	60111695.F1	NIH_MGC_16	Homo sapiens cDNA clone	IMAGE:3357421	5'					
QY	1025	CCGAGAGACTCCGGCTCCGGGTGCGAGGGGAGGAGAAACACAGATGTGGAGATCTCAC	BE253049	544 bp	mRNA	EST	13-JUL-2000					
Db	313	CCGAGAGACTCCGGCTCCGGGTGCGAGGGGAGGAGAAACACAGATGTGGAGATCTCAC	60111695.F1	NIH_MGC_16	Homo sapiens cDNA clone	IMAGE:3357421	5'					
QY	1085	TGTCTCCGATTTCTTCCTTTAAAGGTTCTCATGTCACTATGAGAGGACATGGAGCTCT	BE253049	544 bp	mRNA	EST	13-JUL-2000					
Db	373	TGTCTCCGATTTCTTCCTTTAAAGGTTCTCATGTCACTATGAGAGGACATGGAGCTCT	60111695.F1	NIH_MGC_16	Homo sapiens cDNA clone	IMAGE:3357421	5'					
QY	1145	CTGGACACAAAGCTCTCTCTCTCTTTGATGTGAGACAAAGCTTTCAGGAGAGAGCTGCAG	BE253049	544 bp	mRNA	EST	13-JUL-2000					
Db	433	CTGGACACAAAGCTCTCTCTCTCTTTGATGTGAGACAAAGCTTTCAGGAGAGAGCTGCAG	60111695.F1	NIH_MGC_16	Homo sapiens cDNA clone	IMAGE:3357421	5'					
QY	1205	CTGATCTGGGCTCTGGAATCCGAGATCTCATCGAAGTCTGGGGCTGAAGCTCTCACCTG	BE253049	544 bp	mRNA	EST	13-JUL-2000					
Db	493	CTGATCTGGGCTCTGGAATCCGAGATCTCATCGAAGTCTGGGGCTGAAGCTCTCACCTG	60111695.F1	NIH_MGC_16	Homo sapiens cDNA clone	IMAGE:3357421	5'					
QY	1265	TTTCGAGCCAAAGCCAAACATGTGAGACAAATAGCTCCCAATTTATTTATTTGTGATTTTC	BE253049	544 bp	mRNA	EST	13-JUL-2000					
Db	553	TTTCGAGCCAAAGCCAAACATGTGAGACAAATAGCTCCCAATTTATTTATTTGTGATTTTC	60111695.F1	NIH_MGC_16	Homo sapiens cDNA clone	IMAGE:3357421	5'					
QY	1325	GCCCATTAAGGGCTAACAGAACTGAATTAAGTCTGTTT	BE253049	544 bp	mRNA	EST	13-JUL-2000					
Db	600	GCCCATTAAGGGCTAACAGAACTGAATTAAGTCTGTTT	60111695.F1	NIH_MGC_16	Homo sapiens cDNA clone	IMAGE:3357421	5'					

Accession	Version	Keywords	Source	Organism
BE688967	EST.	human.	Homo sapiens	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
BE688967.1	GI:10317730	1 (bases 1 to 904)	NIH-MGC http://mgc.nci.nih.gov/ Unpublished (1999)	National Institutes of Health, Mammalian Gene Collection (MGC)
REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
				Email: cgapbs-rt@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.llnl.gov Plate: LHM9565 row: e column: 08 High quality sequence stop: 589.
FEATURES	source	Location/Qualifiers		
		1..904	/organism="Homo sapiens" /db_xref="taxon:9606" /clone_1="IMAGE:3848815" /clone_lib="NIH-MGC-65" /tissue_type="adenocarcinoma" /lab_host="DH10B (phage-resistant)" /note="Organ: colon; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.8 kb. Library constructed by Life Technologies."	
BASE COUNT	217 a	235 c	251 g	201 t
ORIGIN				
Query Match	Best Local Similarity	16.7%;	Score 325.8;	DB 141; Length 904;
Matches 444;	Conservative	0;	Mismatches 87;	Indels 6; Gaps 6
OY	774	CTCTGACATCTTTACACTCAAGATCCGGGTCCGGGCTGACCTTAGTGAGACTGCCCTGAC	833	
Db	31	CCCCGACATCTTCCTCCACTGAATCGATTCCGGGTCCGGGCTGACCTTAGTGAGA-TGCCCTCAG	89	
OY	834	GATGTCGAGACCCTTCAGATGTGTGTGATCATGACATGAGCAATCATTTGGGTGTCTCC	893	
Db	90	GATGTCGAGACCCTTCAGATGTGTGTGAGCACATGAGCCACACCACTTGGGTGTCTCC	149	
OY	894	AAACAGGATTTCTTTGCTTTTGGAGAGAGTAACGTCTCTACTGCCAACCCCTAGTAC	953	
Db	150	AAGCAGGATTCCTTTTTC-TTTTGGAGAGAGAGATCATCACTACCTGACCTCCAGAGAC	208	
OY	954	CCTAAGCTTGGAGTGGGTGACATCATTTGATTTGTGTGTGCTGTGACAAGCTCTTCAGAGGC	1013	
Db	209	CCTAAGCTTGGAGTGGGTGACATCATTTGATTTGTGTGTGCTGTGACTTGTGTGCTTACAGAGTCTTCAGAGGC	268	
OY	1014	CACAGAGACATCCACAGAGCTCCGGCTCCGGGTGACAGGGAGAGAGAAACACAGATGTT	1073	
Db	269	CACAGAGAGCTCCCAACAGCTCCAGCTCCGGGTGACAGGGAGAGAGAAACAGACACAT	327	
OY	1074	GGAGATCTACAGCTCTTCGATTCCTCTTTAAGGTTTCATGTCAACATATGAGAGAC	1133	
Db	328	GGAGATCTACAGCTCTTCGATTCCTCTTTAAGAG-CATATGTCCACTATGAGAGAGC	386	
OY	1134	CATGGACATCTTGACACAGCTCTCTTTCTTTTGATGGAGACAAGCTTTTCAGACAA	1193	
Db	387	CATGGACATCTTGACACAGCTCTCTTTCTTTTGATGGAGACAAGCTTTTCAGAGC	446	
OY	1194	GGAGCTGCACGATGATCTGGGCTCGAATCCGAGATCTCATGGAAGTCTGGGGCTGAAG	1253	
Db	447	GGAGCTGCACGATGATCTGGGCTCGAATCTGGGAGACATCATGAGGCTGGGGCTGACA	505	

QY	1254	CHCPCACCCCTGTTGGAGCAAGCCAGACATGTGAGACAATAGCTCCCAATTTAT	1310
Db	506	ACCCATCACTGTT-TGAGGCGCACGCTTGACTTTGGGAGAAATGACTTTCCTTTT	561
RESULT	8		
LOCUS	BC026071	799 bp	EST
DEFINITION	6022922631 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4387022 5'		
ACCESSION	BC026071		
VERSION	BC026071.1	GI:12413308	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
TITLE	1 (bases 1 to 799)		
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/ .		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMN) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/MLN at: http://image.llnl.gov Plate: LHAM10070 row: n column: 15 High quality sequence stop: 529. Location/Qualifiers 1. 799 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4387022" /clone_1ib="NIH_MGC_86" /lissue_type="osteosarcoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: bone; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dr primed. Average insert size 1.533 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH-MGC Library."		
BASE COUNT	210 a 239 c 210 g 140 t		
ORIGIN			
Query Match	15.9%; Score 309.6; DB 172; Length 799;		
Best Local Similarity	74.3%; Pred. No. 4.4e-74;		
Matches	547; Conservative 0; Mismatches 154; Indels 35; Gaps 11;		
QY	641	AGAACTAAGGAGAACTGAACAAGCGCTTCACAGATCTCGCTCTGAGCCCCAAGC	700
Db	2	AGAAGTTAAGTAGGAGAAACAAGCGCTTCACAGATCTCGCTCTGAGCCCCAAGC	61
QY	701	AGCACCAGAGTCCAGCCCTTCAGAGCACAGATGATGAGTGGTCTTAATGGAAGGCGCTG	760
Db	62	CACCTCAGAGGCTTAAAGCAACAGGCGCAAGAGATGAAGTGGTCTTGGGAAGGCGCA	121
QY	761	TCTTCCACAGAGCTCTGACACTTAACTTAACACTCAACATCCGCGCGGCTGACCTAGTGA	820
Db	182	GATTCGCCCTCGAGATGTGAGAGCCCAATGCAAGATGTGTGGACACATGCGCACCCAC	241
QY	880	CTTGGGAGTCTCCAAACAGATTTCTTTTCTTTTGGAGAGAGTGAACGTCTCTACT	939
Db	242	CTTGGGAGTCTCCCAAGCAGATCTCTTTT-TGCTTTTGGAGAACAGACTATCACTACT	299

REFERENCE	Bovidae: Bovinae: Bos.
AUTHORS	1 (bases 1 to 525) Sugimoto, T., Hirotsune, S., Takasuga, A., Itoh, R., Itohzono, A. and Suzuki, H.
TITLE	bovine cDNA sequencing
JOURNAL	Unpublished (2000)
COMMENT	Contact: Yoshikazu Sugimoto Animal Genetics Division Shirakawa Institute of Animal Genetics Ogikura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan Tel: 81-248-25-5641 Fax: 81-248-25-5725 Email: kazusugi@cocoa.ocn.ne.jp Single pass sequencing. This clone was obtained from a polyA-deleted cDNA library.
FEATURES	Location/Qualifiers
Source	1..525
	/organism="Bos taurus" /db_xref="taxon:9913" /clone="E1U035D02" /clone.lib="Bos taurus Lung fetus" /tissue.type="lung" /dev_stage="fetus" /lab_host="DH10B" /note="Vector: pZL1; Site_1: SalI; Site_2: NotI; Poly A was deleted from a NotI site"
BASE COUNT	138 a 144 c 147 g 94 t 2 others
ORIGIN	
Query Match	13.6%; Score 264.2; DB 31; Length 525;
Best Local Similarity	74.0%; Pred. No. 1.2e-61;
Matches 364; Conservative	0; Mismatches 119; Indels 9; Gaps 2
OY	346 CCGGTCGTCCTCCAGTGTACTCCGGGAAGTACAGAGAGCGCTCAACCTCATTCAGATTAAT 405
Db	1 CCGGGGTCCTCCAGTGTACTCCGGGAAGTGAAGAGACGCTTCACCTTATCCAGATAAC 60
OY	406 TCATCCTCTTGAAACTGTGCTCCCTTCAGAGCCTGAAGATGAGGACAGATCTGACAAATTCT 465
Db	61 GTGTCCCTCCGAAATTAATGATCCCGCAGAGGCTGAAGAGAGGACAGATGTGGAGATGCC 120
OY	466 GGCAGTTCCTCCGAGGATGATGCCCTTCAGTTCCTCCGGAAGAAAGAGCTC 525
Db	121 AGCAG--TCCCATGCACTGATGATTTCCATGTCCTCCCTGGAAGAGAGCT- 176
OY	526 AGAAGAAGTGTGAGAAAGAAAGAAAGAAATGAGAAGTTCCTCCGACAGACATCTCT 585
Db	177 ----GAGGAGTAAGGATGAGAGAGAGAAAGAAAGATTTTACTGATCAGGACACTCG 231
OY	586 CCTTTGGCCCAACCTTCGTCAAGGAACAAGACAGAAACATACGAGGCGCTCCAGANG 645
Db	232 CCTTTGCCCTCACTCTGTCAGCTACCAAAACAGAGAACATACAGGACACTCCGGAAG 291
OY	646 CTAAAGGAAGTAAACAAGGCTCCCAAGATTCGGCTCTGGCTGGGCCCAAGACAGCAC 705
Db	292 TTAAGGAGAGTAAACAAGGCTCTCCAGATTCCTTTCTGCTGATGCCAAGACATCC 351
OY	706 CAGAGTCAGCCCTTCAGACACAGATGATGAGTGTGCTCTAGTGAAGAGCCCTGTCTTG 765
Db	352 CAAGCCCAAGACACACTGTCAGCCACAGATGAGTGTGCTCTAGTGAAGAGGCCCAAGNC 411
OY	766 CCACGAGAGTCTGCAGCTTTTACAGTCAAGATCCGGTGGCGGGCTGACTAGTGAACAG 825
Db	412 CCAGGAACTCTGCAGTCTCTGCACACTCAAAAATCCGGTGGCGAGTGAAGTCAAGTTG 471
OY	826 CCTGTCAGGATG 837
Db	472 CCCATCAGAAATG 483
RESULT 11	
BF310650	

LOCUS	BF310650	669 bp	mRNA	EST	21-NOV-2000
DEFINITION	60189460552 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4124232 5', mRNA sequence.				
ACCESSION	BF310650				
VERSION	BF310650.1 GI:11258236				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 669)				
TITLE	NIH-MGC http://mgc.nci.nih.gov/.				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cga@bbs-remail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Ling Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: image.lnl.gov Plate: LNCMI010 row: 1 column: 01 High quality sequence stop: 600. Location/Qualifiers 1..669 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4124232" /clone_lib="NIH_MGC_19" /tissue_type="neuroblastoma" /lab_host="DH10B (phage-resistant)" /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC library."				
FEATURES	source				
BASE COUNT	174 a 191 c 184 g 120 t				
ORIGIN					
Query Match	13.3%; Score 258.8; DB 147; Length 669;				
Best Local Similarity	70.4%; Pred. No. 4.1e-60;				
Matches 432; Conservative	0; Mismatches 172; Indels 10; Gaps 6;				
416	TGAACCTGGCCCTTCACAGGCGCTGAAAGTGAAGGACATCTGCAAAATCTGSCAGTTCTC	475			
2	TGAAACTCTACCTCCACG-GGATGAGGAAGAGGACGCTGGCAGATTTC--GAGTGGTC	57			
476	CCTCGAGGATGATCCGCTCGCTTCAAGTTCCCTCGAGAAAGAAAGCTCAGAAAGAGT	535			
58	TCTACCATGAGGGCTCCCATCACCAGGGTCCCTGGGAAGCAAAAGCTGAGAGCTAAG-	116			
536	GTTGAAGAAGAGAAAAAGAAATGTAAGAGTTTCGAGCACCAGACATCTCTCTTGGCCC	595			
117	--GATTAAGAAAGAGAAAAAGAAAGAGATTCTCGATCTTGACAACTCTCTCTGTCC	174			
596	AACCTTGTGCAAGAAACAAAGCAGAAAGCATATGAGGAGCGCTCCAGAAAGCTAAAGGAAG	655			
175	CACCTTCAACAAAGAGCAAAAGCAGACGACATCTCGGCACTCAAGAAAGTTAAATGAGC	234			
656	TGAACAGCGTCTCCAGATCTCGCTCTCGCTGAGCGCCCAAGCAGCAGACGATCCAG	715			
235	TGAACAGGCGCTCCAGGATCTCGCTTCTGCTGAGCGCCCAAGCCACCTCAGGGTCAAG	294			
716	CCCTTCAGAGCAGCAATATATGAGTGGTGGTCCATAGTGAAG-GGGCGTGTCTTGCCACAGC	774			
295	AGCAACAGGGGCCAAGAGATGAGTGGTGTGGTGAAGGGGCCAGTCCGCAAGAGAGC	354			
775	TCTGCACCTTTACACTCAAGATCCGGTGCCCGGGGTGAACCTGTAGAGACTGCTTCAGG	834			

FEATURES

	source	1. 239 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="4930580L19" /clone_id="RIKEN full-length enriched, adult male testis (DH10B)" /sex="male" /tissue_type="testis" /dev_stage="adult" /lab_host="DH10B" /note="Site 1: Salt; Site 2: BamHI. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGGAGGAGGAGATCCAAAGCTCTTTTTTTTTTTTTTVN 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adaptor of sequence [5', GAGGAGGAGATTCTCGACTGTATTAATAATGCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pluscript KS(+) after bulk excision from LambdaFLC I. Cloning sites, 5' end: Salt; 3' end: BamHI."
BASE COUNT	63 a 53 c 56 g 67 t	
ORIGIN		
Query Match	11.5%; Score 224.6; DB 123; Length 239;	
Best Local Similarity	96.2%; Pred. No. 7.7e-51;	
Matches 230; Conservative	0; Mismatches 9; Indels 0; Gaps 0;	
QY 1115	TGTCACACTATGAGAAGCATGGAGCACTCTCGACACAAGGCTCTCTTCTTGATG 1174	
Dd 1	TGTCACACTATGAGAAGCATGGAGCACTCTCGACACAAGTTCCTTTCTTGATG 60	
QY 1175	GGACAAAGCTTTCAGGCAGAGAGCTGCCAGCTGATCTGGGCTGTGAATCCGAGATCTCA 1234	
Dd 61	GGACAAAGCTTTCAGGCAGAGAGCTGCCAGCTGATCTGGGCTGTGAATCCGAAATTTTCA 120	
QY 1235	TGGAGTCTGGGGGCTGAAGCTCTCACCCCTGTCGGACGCAAGCCAAAGCATGGAGACAA 1294	
Dd 121	TGCAAGTTTGGGGCTGAAGCTCTCACCCCTGTCGGACGCAAGCCAAAGCATGGAGACAA 180	
QY 1295	TAGCTCCCAATTTTATTTATGTGATTTTTCGCCCATTAAGGCTTAACGAATCTGAATT 1353	
Dd 181	TAGCTCCCAATTTTATTTATGTGATTTTTCGCCCATTAAGGCTTAACGAATCTGAATT 239	
RESULT 14		
BG392671	616 bp mRNA EST 12-MAR-2001	
LOCUS	602410920R1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4539700 5',	
DEFINITION	mRNA sequence.	
ACCESSION	BG392671	
VERSION	BG392671.1 GI:13286119	
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 616) NIH-MGC http://mgc.nci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)	
AUTHORS	Contact: Robert Strausberg, Ph.D. Email: csapbs@mail.nih.gov	
JOURNAL	Tissue Procurement: ATCC	
COMMENT	cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)	

DNA Sequencing by:	Incyte Genomics, Inc.
Clone distribution:	MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.jnl.gov
Plate:	LlAM10468 row: h column: 05
High quality sequence stop:	612.
FEATURES	
Source	Location/Qualifiers 1..616 <code>/organism="Homo sapiens"</code> <code>/db_xref="taxon:9606"</code> <code>/clone_image="IMAGE:4539700"</code> <code>/clone_lib="NIH_MGC_92"</code> <code>/tissue_type="embryonal carcinoma, cell line"</code> <code>/lab_host="DH10B (phage-resistant)"</code> <code>(note="Organ: testis; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: This is a NIH-MGC library.")</code>
BASE COUNT	117 A 176 C 228 G 95 T
ORIGIN	
Query Match	11.2% Score 218.2 DB 153 Length 616;
Best Local Similarity	67.5%; Pred No. 6e-49; Mismatches 163; Indels 25; Gaps 5;
Matches 391: Conservative	0; Mistmaches 163; Indels 25; Gaps 5;
OY	48 CGCGGTGGCCAGGACGCTCAGAGAACCAGAGCCCCTGTCGGTGCTCTCGGCCG 107 DB 44 CGGTCCGGCCGCAAGGGGGGTGGGGGGCGTTGGGGCGGTGCGGGCGCTCCGGGGCCA 103
OY	108 GCATGCTCCGGCTAGCGCTCATTCACAACACCGTGCTTGTGAATTGGTCACTAAGACGGA 167 DB 104 GGCGTCTCATCCCCGGGGCACGCTGGACGTACTGTCTGTGGACTTGGTCACCGACGCA 163
OY	168 CGAAGAGGCTTTGSAGT-----CGCAGACCGAGTAGAGGTGCGGT 209 DB 164 TGAGAAATAATTCTGGAGGTGCGCACCGCTGCGGGTCCCGCGAGAGTTGAGGTGAGGC 223
OY	210 CGCCCCCTCCCCCGCGGAGCTTAACCTGAGGAGGACAAGCAGCTGACAGTGAAGAGGGGC 269 DB 224 CCCGGAGCCCCCGGGGGCGGGTGGCGTCCCGGGGATMAACAGCAAATGACAGCGAAGGGGA 283
OY	270 GCCCGAGGGGCTCGCGGAGACCCCGCTACATTTGTGCGACGCGGGCGGGCGCTGCT 329 DB 284 GGACAGGGGGCCCGCAGAGGCCCGCGGGAGCGGTCAAGCGCGCGCTGTGCT 343
OY	330 GGATCCCGAGAGGGCGGTGGTCCAGTGTACTCCGGGAAGGTACAGAGCAGCTCAA 389 DB 344 GGATCCGGGGAGGGCGCGCTGCTGGTGGTACGTGGGGGAAGTTTTAAAGAGAGCTTCG 403
OY	390 CCTCATTCAGATTAATTCATCCCTCTTGAATGTCCTTCAGACCTTGAAAGATGAGGC 449 DB 404 CCTTATCCAGATGATGTATCCCTCTCTGAACATTCATCCCTCAG -GATAGAGGAAGAGGC 462
OY	450 AGATCTGACAAATTCGTGCGAGTTCTCCCTCTGAGAGATGATGCCCTTCAGGTTCTCC 509 DB 463 AGAGCTGGACAGATTC ---GAGTGTCTCTACCATGAGGGCTCCCATCCAGCGCTTCC 519
OY	510 CTGGAGAAAAGAGCTCAGAAAAAAGTGTGAGAAAAGAAAAGAAAATGGAGAGATTTCC 569 DB 520 CTGGAGAACAAAGCTCAGAGACTTAAAGAT--AAACCAAGAGAAAAGAAAGACAGAGATTCT 577
OY	570 GGACGAGACATCTCTCTTTGCCCAACATCTGTCAG 608 DB 578 GGATCTGGACA ATCTCCTCTGTCTCCCACTTCACCAAG 615
RESULT 15	
BFO76210	BF076210 544 bp mRNA EST 18-Oct-2000
LOCUS	DEFINITION 225720 MARC ZBOV Bos taurus cDNA 5' , mRNA sequence.
ACCESSION	BF076210

VERSION BF076210.1 GI:10869849

KEYWORDS EST.

SOURCE COW.

ORGANISM

Bos taurus

Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovidae; Bovinae; Bos.

1 (bases 1 to 544)

Smith, T. P. L., Casas, E., Stone, R. T., Heaton, M. P., Grosse, W. M.,

Bennett, G. A., Fahrenkrug, S. C., Freking, B. A., Rohrer, G. A., Laegreid

W. W. and Keele, J. W.

Design and use of four pooled tissue normalized cDNA libraries for

EST discovery in cattle

Unpublished (2000)

Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt trimmed with phred

v0.980904.e. Vector identified by cross_match with the -mnscore 18

and -mismatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAAT

BACKWARD: GTTTCAGTCAGCAGC

Plate: 86 row: G column: 5

Seq primer: ATTAGCTGACACTATAG.

Location/Qualifiers

1. 544

/organism="Bos taurus"

/db_xref="taxon:9913"

/clone_1bp="MARC 280Y"

/tissue_type="pooled"

/lab_host="DH10B"

/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;

Library made from pooled tissue from testis, thymus,

semilendonsus muscle, longissimus muscle, pancreas,

adrenal, and endometrium."

BASE COUNT 106 a 156 c 147 g 135 t

ORIGIN

Query Match 10.6%; Score 206.6; DB 144; Length 544;

Best Local Similarity 79.8%; Pred. No. 9.1e-46;

Matches 256; Conservative 0; Mismatches 64; Indels 1; Gaps 1;

QY 978 CATTGATTGTGTGCTAGCAGCTCTTTCAGAGCCACAGACATCCAGAGCTCCG 1037

DB 35 CACAGACTGTGTGTAGCAGTCTCCGAGGCTGCAGAGAGCTGCCACTGCTCCA 94

QY 1038 GCTCGGGGTGAGGGGAAGGAGACACACAGATTTGAGATCTCAGTCTCCTGATTC 1097

DB 95 ACTTCGGGTCCAGGGAAGGAGACACAGCTGGAAGTCTCGCTCCCTCCGATTC 154

QY 1098 TCCCTTAAGTTCATCTACACTATGAGAGCCATGAGACTCTGTGACACAAGCT 1157

DB 155 TCCCTCAAGACCTCATGTCCGCTAGAGAGGCTATGAGACTCTCGGCCACAAGCT 214

QY 1158 CTCCTTTCTTTGATGGGAAAGCTTTTCAGGCAAGAGCTGCCAGCTGATCTGGGCT 1217

DB 215 CTCCTTTCTTTGATGGGAGGAGGAGCTTTTCAGGCAAGAGCTGCCAGCTGAGGAAAT 274

QY 1218 GGAATCCGAGATCTCATGCAAGCTGGGGCTG-AAAGCTCACCCTGTGCGAGCAAA 1276

DB 275 GGAATCTGGGAGACTCATTCGAGGCTTGGGGCTGAACCCCACTCCCTTTGGAGGCCA 334

QY 1277 GCCAAGACATGAGACAATAG 1297

DB 335 GCCTGACTTGGGAGAAATGG 355

Search completed: September 11, 2001, 12:52:44
Job time: 4256 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 11, 2001, 11:44:13 ; Search time 138.91 Seconds
(without alignments)
8796.318 Million cell updates/sec

Title: US-09-192-611-1

Perfect score: 1946
Sequence: 1 ACAGTGTGGAGATGCGCGA.....CGCATTCGCCGACTCGAG 1946

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq_0601:*

1: /SIDSL/gcgdata/geneseq/geneseqn/NA1980.DAT:*

2: /SIDSL/gcgdata/geneseq/geneseqn/NA1981.DAT:*

3: /SIDSL/gcgdata/geneseq/geneseqn/NA1982.DAT:*

4: /SIDSL/gcgdata/geneseq/geneseqn/NA1983.DAT:*

5: /SIDSL/gcgdata/geneseq/geneseqn/NA1984.DAT:*

6: /SIDSL/gcgdata/geneseq/geneseqn/NA1985.DAT:*

7: /SIDSL/gcgdata/geneseq/geneseqn/NA1986.DAT:*

8: /SIDSL/gcgdata/geneseq/geneseqn/NA1987.DAT:*

9: /SIDSL/gcgdata/geneseq/geneseqn/NA1988.DAT:*

10: /SIDSL/gcgdata/geneseq/geneseqn/NA1989.DAT:*

11: /SIDSL/gcgdata/geneseq/geneseqn/NA1990.DAT:*

12: /SIDSL/gcgdata/geneseq/geneseqn/NA1991.DAT:*

13: /SIDSL/gcgdata/geneseq/geneseqn/NA1992.DAT:*

14: /SIDSL/gcgdata/geneseq/geneseqn/NA1993.DAT:*

15: /SIDSL/gcgdata/geneseq/geneseqn/NA1994.DAT:*

16: /SIDSL/gcgdata/geneseq/geneseqn/NA1995.DAT:*

17: /SIDSL/gcgdata/geneseq/geneseqn/NA1996.DAT:*

18: /SIDSL/gcgdata/geneseq/geneseqn/NA1997.DAT:*

19: /SIDSL/gcgdata/geneseq/geneseqn/NA1998.DAT:*

20: /SIDSL/gcgdata/geneseq/geneseqn/NA1999.DAT:*

21: /SIDSL/gcgdata/geneseq/geneseqn/NA2000.DAT:*

22: /SIDSL/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1946	100.0	1946	18	Mouse NF-AT Inter
2	741.4	38.1	2576	20	Human NIP45 CDNA.
3	721.6	37.1	1260	20	Human NIP45 coding
4	197.6	10.2	300	21	Human NIP45 cancer
5	197.6	10.2	300	21	Human NIP45 cancer
6	167.6	8.6	705	21	Human colon cancer
7	46.4	2.4	8438	15	Human colon cancer
8	44	2.4	5232	13	AAQ73500
9	42.8	2.2	114955	20	AAQ31891
10	41.8	2.1	114955	20	AAQ31891
11	41.6	2.1	700	21	Human adenosine A1
					Human adenosine A1
					FMRI gene variable

C	12	41.6	2.1	2183	13	AAQ28290	Fragment from huma
C	13	40.2	2.1	3656	20	AAQ23299	Human SPA-1 CDNA.
C	14	40.2	2.1	4425	18	AAQ72781	Human interleukin-
C	15	39	2.0	2375	21	AAQ20862	Human beta-2 adren
C	16	39	2.0	2375	21	AAQ34740	Human adenosine re
C	17	39	2.0	4475	21	AAQ20863	Human beta-2 adren
C	18	39	2.0	4475	21	AAQ34741	Human adenosine re
C	19	39	2.0	15872	18	AAQ68715	Streptomyces venez
C	20	39	2.0	35384	21	AAQ21436	Human enzyme-relat
C	21	39	2.0	58857	21	AAQ58471	Nucleotide sequenc
C	22	38.6	2.0	237326	19	AAQ57903	Hereditary haemoch
C	23	38.4	2.0	4632	22	AAQ54902	Human activity dep
C	24	38.4	2.0	235033	19	AAQ57926	Hereditary haemoch
C	25	38.2	2.0	4100	20	AAQ32019	Human MERTH relate
C	26	38.2	2.0	4100	22	AAQ90076	Human MERTH clone.
C	27	38.2	2.0	4897	11	AAQ03259	Pseudorabies virus
C	28	38	2.0	117213	19	AAQ62176	HSV-2 strain SB5 C
C	29	37.8	1.9	1944	20	AAQ33159	Rhodobacter sphaer
C	30	37.8	1.9	2846	19	AAQ41260	Chlamydomonas rein
C	31	37.8	1.9	3256	20	AAQ99971	Nucleotide sequenc
C	32	37.6	1.9	537	21	AAQ37410	Arabidopsis thalia
C	33	37.4	1.9	1955	22	AAQ31254	Chlamydia pneumoni
C	34	37.4	1.9	15872	21	AAQ87283	S. venezuelae vep
C	35	37.4	1.9	43280	18	AAQ80413	Tyactone synthase
C	36	37.4	1.9	273254	21	AAQ81914	Chlamydia pneumoni
C	37	37.4	1.9	1230025	20	AAQ91990	Nucleotide sequenc
C	38	37.2	1.9	10732	21	AAQ10594	Gene encoding a su
C	39	37.2	1.9	49999	20	AAQ23904	Human LOB homolog
C	40	37	1.9	1002	20	AAQ24498	Human SR-BI gene p
C	41	37	1.9	1002	20	AAQ24562	Human SR-BI gene p
C	42	37	1.9	1002	20	AAQ24590	Human SR-BI gene p
C	43	37	1.9	1560	21	AAQ24654	Human SR-BI gene p
C	44	37	1.9	1560	21	AAQ34962	Corn neutral triac
C	45	37	1.9	29879	14	AAQ46806	erya region of S.

ALIGNMENTS

RESULT	1
AAQ93045	standard; CDNA; 1946 BP.
ID	AAQ93045
AC	AAQ93045;
XX	
DT	18-MAY-1998 (first entry)
XX	
DE	Mouse NF-AT Interacting Protein 45 coding sequence.
XX	
XX	NF-AT Interacting Protein 45; NIP45; yeast two-hybrid assay; mouse;
KW	Rel homology domain; RHD; T cell; transcription factor; cancer;
KW	interleukin-4; IL-4; development; Th1; Th2; cytokine; allergy;
KW	autoimmune disease; transplantation; ds.
OS	Mus sp.
XX	
XX	
FH	Key
FT	13.1251
FT	Location/Qualifiers
FT	/*tag= a
FT	/product= NIP45
XX	
XX	WO9739721-A2.
PN	
PD	30-OCT-1997.
XX	
XX	23-APR-1997;
PE	97WO-US06708.
XX	
XX	25-NOV-1996;
PR	96US-0755592.
PR	23-APR-1996;
PR	96US-0636602.
PR	25-NOV-1996;
XX	96US-0755584.
XX	
XX	(HARD) HARVARD COLLEGE.
XX	

PI Glimcher LH, Ho I, Hodge MR;
XX WPI; 1997-535556/49.
XX Production of cytokine(s) associated with Th2-type helper T cells -
PT particularly for controlling development of Th1 and Th2 cells for
PT treatment of allergy, autoimmune disease etc.

PS Claim 40; Fig 11; 151pp; English.

XX This sequence encodes the mouse NF-AT interacting protein 45 (NIP45).
CC The sequence was isolated by using a yeast two-hybrid detection system
CC for proteins that interact with the NF-AT Rel homology domain (RHD).
CC The assay used, as a "bait", a 900 bp fragment of the murine NF-ATP
CC encoding the region spanning amino acids 228-250. NF-AT is a
CC multisubunit transcription complex containing a cyclosporin A-sensitive
CC cytoplasmic phosphoprotein and an inducible component of the AP-1 family
CC of transcription factors. The screen was carried out on a cDNA library
CC prepared from the murine T cell line D10. One class of proteins, NIP45
CC designated NIP45, bound the NF-AT-RHD region with high affinity. NIP45
CC can be used to induce NF-AT expression and can be used in a claimed
CC method to inhibit or stimulate production of NF-AT family proteins,
CC particularly interleukin-4 (IL-4). Cells treated in such a manner may be
CC administered to control development of Th1 or Th2 cells by modulating
CC the production of a T-helper type 2 associated cytokine. The method is
CC especially useful in the inhibition of Th2 in allergy, cancer or
CC infections, and promotion of Th2 in autoimmune disease and
CC transplantation.

XX Sequence 1946 BP; 438 A; 500 C; 562 G; 446 T; 0 other;

Query Match 100.0%; Score 1946; DB 18; Length 1946;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1946; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAGTGGGAGATGGGGGAACCACTAGGGGAGCTGTCCGAGTCCCGGTGGCCGA 60
DB 1 acagtgggagatgggggaacccactagggagcgtgtcccgagtcgccggtggccga 60
QY GGGGCTGGAGAGCCCGAGCGCCGCGGCTGCTCGCGCGCGGAGTCCCGGCT 120
DB ggggctggagagcccgagcgccgctgctccgctcccgcccgagtcgccgct 120
QY 121 AGGCTATTCCAGACACCGCTGCTGTGGACTTGTCAGTACAGCAGAAAGTCTTG 180
DB 121 aggctattccagacacccgctgctgtgacttgctcagtcagcagcgagaggtcttg 180
QY 181 GAAGTCGACACCACTAGTAGAGTCCGGTGCCTCCCGCGCGCTAAACCTGAG 240
DB 181 gaagtcgacacccactagtagagtgctgcgctcccgctcccgcgctaaacctgag 240
QY 241 CAGGACAGCAGATGACATGAAAGGGGCGCGAGGGGCTGGGGAGCCCGCGTACA 300
DB 241 caggacagcagatgacatgaaaggcgcgaggggctggggagcccgcgctacaca 300
QY 301 TTGCTGGAGCGCGCGCGCGCTGCTGATCCCGAGAGGGCGCGGTGCCAGTG 360
DB 301 ttgctggagcgcgcgcgcgctgctgatcccgagagggcgcggtgccagtg 360
QY 361 TACTCGGGGAAGTAGAGAGCAGCTCACTATTCCAATAATTATCCCTTTAAA 420
DB 361 tactcggggaagtagagagcagctcaactatccaataatcctcctttaa 420
QY 421 CTGTGCGCTTCAGAGCTTGAAGATGAGGAGATCTGCAATTTCTGGCAGTTCCCTCT 480
DB 421 ctgtgcgcttcagagcttgaagatgaggagatctgcaaatcttgagcttccctct 480
QY 481 GAGGATGATCCCTGCTTACGTTTCCTCGAAGAAAGAGCTCAAGAAAGAGTGAG 540
DB 481 gaggatgatccctgcttaccgttccctcgtagaagaagagctcaagaagagtgag 540
QY 541 AAAGAAGAAAGAAATGAGAGATTTCGGACCGAGCATCTCTCTTTGCCCAACT 600

DB 541 aagaagaaagaaatgagagattccgaccagacatctctcttgcaccaact 600
QY 601 TCGTCAAGGAGCAAAAAGCAGAAAGCATATGAGAGCGCTCCAGAGCTAAAGGAGTGAAC 660
DB 601 tcgtcaaggagcaaaaagcagaaagcatatgagagcgctccagagctaaaggagtgaaac 660
QY 661 AAGCGTCTCCAGATCTCCGCTCTGCTGAGGCCCAAGCAGCAGCAGATCCAGCCCTT 720
DB 661 aagcgctctccagatctccgctctgctgtaggcccaagagcagagtcagccctt 720
QY 721 CAGAGCAGAGATATGATGCTGCTTAGTGGAAGGGCTTCTTGGCAGAGACTTCCA 780
DB 721 cagagcagagatgctgcttagtggaagggtcttggcagagacttcca 780
QY 781 CTTCTTACATCAAGATCCGCTCCGGGCTGAGCTAGTGAAGTGGCTGACAGATGTG 840
DB 781 ctcttaccatcaagatccgctccgggctgagctagtgagctgctgcagagtcg 840
QY 841 GAGCCCTTCAGAAATGTGTGATCAATGCGCAATCATCTTGGGTGCTCCAAACAG 900
DB 841 gagcccttcagaaatgtgtgatcaatgcgcaatcatcttgggtgctccaacag 900
QY 901 ATTCTTTGCTTTTGGAGAGAGTGAACGTGTCTTACTGCAACCCCTAATACCTTAAG 960
DB 901 attctttgcttttggagagagtgaacgtgtcttactgcaacccctaataccttaag 960
QY 961 CTTGAGTGGCTGACATCATTTGATTTGTGTGCTAGCAAGCTCTTCAGAGGCGCACAG 1020
DB 961 cttgagtggctgacatcatttgatTTGTGTGCTAGCAAGCTCTTCAGAGGCGCACAG 1020
QY 1021 ACATCCAGAGAGCTCCGGCTCCGGGTGCAGGGAGAGCAACACAGATGTTGAGATC 1080
DB 1021 acatccagagagctccggctccgggtgcagggagagcaaacagatgTTGAGATC 1080
QY 1081 TCACGTGCTCCTGATTCCTCTTAAGGTTCATGTACACATATAGAAAGCATAGGA 1140
DB 1081 tcacgtgctcctgatTCCTCTTAAGGTTCATGTACACATATAGAAAGCATAGGA 1140
QY 1141 CTCTGTCGACACAAGCTCTCTCTTTTGTATGGAGCAAAAGCTTTCAGGACAGAGCTG 1200
DB 1141 ctctgtcgcacacaagctctctcttttgtatggagcaaaagcttccagagcaagagctg 1200
QY 1201 CCACTGATCTGGGCTCGAATCCGAGATCTCATGCAAGCTGGGGCTGAAGCTCTCAC 1260
DB 1201 ccactgatctgggctcgaatccgagatctcatgcaagctggggctgaagctctcac 1260
QY 1261 CCTGTTGGAGCGCAAGGCCAAGACATGAGACAAATAGCTCCCAATTTTATTTGTGATT 1320
DB 1261 cctgttggagcgcaagggccaagacatgagacaaatagctcccaatttatattgtgatt 1320
QY 1321 TTTGCGCCCAATAAGGGGCTACAGAAACTGAATTAGAACTGTTTACTTATTTATCTGG 1380
DB 1321 ttgctgggattgaacccagacatgacatgcaatgaaactgttacttatattcttcg 1380
QY 1381 TGCTGGGATGGAACCCAGACATGACATGCTTAAGAGATGTATGAGTGGAGCAAAAC 1440
DB 1381 tgctgggattgaacccagacatgacatgcttaagagatgtatgagtgaggcaaaac 1440
QY 1441 CAAGGATTAACCTTTAGCCGCTTAGTAGACTGTAGTCAAGCAAGTGGCTACTTG 1500
DB 1441 caaggattaacctttagccgcttagtagactgttagtcaagaagtggctacttg 1500
QY 1501 TAGTGTGCTGCTGCTGTATGTTTGTGCTGATTTGGAGCGCCCTGGGCGACATGAAG 1560
DB 1501 tagtgtgctgctgctgtatgTTTGTGCTGATTTGGAGCGCCCTGGGCGACATGAAG 1560
QY 1561 GGAAGCTTGGCTTCCATACATTTACAGTTTGCTGTGCTCTTCTTCATCAGATGACTT 1620
DB 1561 ggaagcttggcttcccatcacatTTTACAGTTTGCTGTGCTCTTCTTCATCAGATGACTT 1620
QY 1621 CTGTGAAGCTGCTTATGTTAGTGTGTGTAACATAATGAGCTTGCTTTGGGTGCTCAG 1680

Db 1621 ctgtgaagctgtcctatgttgatgtgtgtgaactaaatgagctcgttgggtgccag 1680
 QY 1681 CCTGGGCTTTTGGCCGAGTGGAGCCAGACGTGACTTGCATTGGAGACTGAGAA 1740
 Db 1681 cctgggcttttggccgagcttgagccagcagtgtaactcctcgtactggactgagaa 1740
 QY 1741 TGCATTTCTCTGTTGGAGACACTCGGTGACAGAAATATTAACAGAGTGACATACGCTG 1800
 Db 1741 tgcattcctgtgtgagacactcgggtgcagaataataacagaagtgatcatatgctgt 1800
 QY 1801 AAGCTGAGGACTAGTGGAAAGTTAAGACGCTTGCTATTTTACGCTTGGATCTCTCT 1860
 Db 1801 aagctgagagactagtcgaaagttacagcgttgcatttgcgttggatcctcctct 1860
 QY 1861 GCTGCAGACGACTAGCTGTGTGTGTATACACTTCTTGACATGGACACTGAGTGA 1920
 Db 1861 gctgcagagactagctagcagtgctcgtgtacacactcttgcgtatgacactagtcga 1920
 QY 1921 CGCGGGCGCGATTGCGCGACTCGAG 1946
 Db 1921 cgcgggcgcgactcgcgcgactcgag 1946

RESULT 2

AA56954
 ID AAX56954 standard; cDNA; 2576 BP.

AC AAX56954;

DT 16-JUL-1999 (first entry)

XX Human NIP45 cDNA.

XX NIP45, human: transcription activator; IL-4; interleukin-4;
 KW NF-AT interacting protein-45; modulator; treatment; inflammation;
 KW autoimmune disease; HTP screening; drug testing; allergic therapy;
 KW T-cell dependent response; cytokine; diagnostic; immunosuppressant; ss.

OS Homo sapiens.

PN W09921993-A1.

PD 06-MAY-1999.

PF 21-OCT-1998; 98WO-GB03141.

PR 24-OCT-1997; 97GB-0022388.

PA (ZENEC) ZENECAL LTD.

PI Liu D, Zhao J, Zhou H;

DR WPI; 1999-312964/26.

PS Polynucleotides encoding human NF-AT interacting protein (NIP45)

XX Example 1; Fig 1; 84pp; English.

CC This invention describes a novel human NF-AT interacting protein-45
 CC NIP-45. Human NIP45 is a transcriptional trans-activation factor of the
 CC Interleukin 4 (IL-4) gene. Modulators of human NIP45 and IL-4 can be used
 CC to treat conditions mediated by NIP45 or IL-4, e.g. inflammation or
 CC autoimmune disease. NIP45 is a good candidate target for HTP screening
 CC and/or testing system for drugs that will alleviate T-cell dependent
 CC autoimmune and allergic responses, and for cytokine-based therapies of
 CC chronic disease. Expression of NIP45 can be inhibited, and IL-4
 CC expression modulated, by administering an antisense molecule. Antibodies
 CC against NIP45 and primers can be used in diagnostic assays. Discovery of
 CC a human NIP45 is advantageous in that it provides the ability to control
 CC IL-4 transcription, which is of importance for anti-inflammation and
 CC immunosuppressant drug development.

SQ Sequence 2576 BP; 549 A; 689 C; 710 G; 628 T; 0 other;

Query Match 38.1%; Score 741.4; DB 20; Length 2576;
 Best Local Similarity 74.5%; Pred. No. 9,4e-198;
 Matches 1033; Conservative 0; Mismatches 306; Indels 48; Gaps 6;
 QY 6 GTGGAGATGGCGGAGAACACATGAGGAGCATGTGTCGAGTCCGCGT-----GG 56
 Db 6 gtggtacatgycgagacctgtgagggaagcgggctggtcgcgaggttagcgtgtccgg 65
 QY 57 CCGAGCGCTGTGAGAGCGCCGAGCGCCGTCGCGGTGCTCTGCGCCCGGACGTCTCC 116
 Db 66 ccgaggggtgtggtggcgtgtggtggtggtggtggtggtggtggtggtggtggtggt 125
 QY 117 GGCATGCTCATTCACAGACACCGCTGTGACCTGTGCTGCTGCTGCTGCTGCTGCTGCT 176
 Db 126 atccggggacacgctgtgagctagtgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 185
 QY 177 CTGGAACT-----CGACACCCAGTAGAGTGGCGGTGCGCCGCT 218
 Db 186 tctgaggttgcacacgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgt 245
 QY 219 CCCCAGCCCGCTTAACCTGACAGAGACGACAGTACAGTAAAGGGCGCGGAGG 278
 Db 246 ccggggcggt 305
 QY 279 GCTGCGGGAGCGCCGCGTACATTGGTGTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 338
 Db 306 gcccagagaccccgcggtggt 365
 QY 339 AGAGCGCGGCTGTGCTCCAGTGTACTCCGAGAGGTACAGACAGCCCTAATCTATTC 398
 Db 366 ggaagcgctgt 425
 QY 399 AGATATTCATCCCTCTTGAACATGTGCTTCAAGAGCTTAAGATGAGGAGATCTGAC 458
 Db 426 agatgtatcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 485
 QY 459 AATTTCTGCGAGTTCTCCCTCTGAGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 518
 Db 486 agattc---gagtggtctctacatcagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 542
 QY 519 GAAGCTCAGAAAGAGTGTGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 578
 Db 543 aaagctgtgagtaag---gataaagaagaagaagaagaagaagaagaagaagaagaaga 599
 QY 579 CATCTCTCTTGTGCCCCCACTTGTGTCAAGAAACAAAGCAAGCAAGCAAGCAAGCAAG 638
 Db 600 caactctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 659
 QY 639 CCAGAGCTAAGGAGAGTGAACAGAGCGTCTCAAGATCTCGCTCCTGCTGAGCCCA 698
 Db 660 caagaagttgaagtgaaggaagaagcgtcctcagagatcctcgtctcgtctcgtcgtcgtc 719
 QY 699 GCAGCACCAGATCCAGCCCTTCAAGACACAGATGATGAGTGTCTTGTGGAAGGCGC 758
 Db 720 gccacctcaggttcaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 779
 QY 759 TGTCTTGCACAGAGCTGTGACCTTTTACACTAAGATCCGGTCCGGGCTGACTAGT 818
 Db 780 caccctcaggaagaccccccagctcctcctcctcctcctcctcctcctcctcctcctcct 839
 QY 819 GAGACTGCTGTGAGATGTGAGAGCCCTTCAAGATGTGTGATGATGATGATGATGATGAT 878
 Db 840 cagatgtcctcctcagatgt 899
 QY 879 TCTTGGGGTGTCTCAAAACAGATCTTTTGTGAGAGTGAAGTGTCTCTTAC 938
 Db 900 ccttggt 959
 QY 939 TGCACCCCTAGTACCTTAAGCTTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 998
 Db 960 tgcacctccagagaccctaaagctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1019

QY	999	AAAGTCCTTCAGAGGCCACAGACATCCAGAGACTCCGGCTCCGGGTGCAGGGGAAGGA	1058
Db	1020	aaagttctccagagcgacacagagacgftcccaacagctccagctcgggtgcagggaaaga	1079
QY	1059	GAAACACCGAGTGTGGAGATCTCACTGCTCTGATTTCTCTCTTAAGGTTCTCATGTC	1118
Db	1080	gaacaaccagacacggaagctcactgctcgcagatctccctctaaagacctcatgtc	1139
QY	1119	ACACTATGAGGAAGCCATGGAGCTCTCTGGACACAAGACTCTCTCTTTGATGGAGAC	1178
Db	1140	ccacatcagagagcgccatcagctcgcagagcgaagacctctctctcttttgatggagcc	1199
QY	1179	AAAGCTTTCAGCGCAAGAGAGCTCCACGCTGATCTGGGCTCGAATCCGAGATCTCATGCA	1238
Db	1200	aaagcttctcagcagggagcctgcagctcagctcagctcagctcagctcagctcagctcag	1259
QY	1239	AGTCGGGGGGCTGA-AGCTCTCAACCTGCTGGAGCCGAAGCCGAAGACATGAGACAAATG	1297
QY	1298	CTCCCAATTTTATTTATTTGATTTTTCGCCCATTAAGGCTTAACAGAACTGATTAAGAA	1357
Db	1320	c-----ttcccttttttggcccataagggctcagctcagctcagctcagctcagctcag	1365
QY	1358	CTTGCTTT 1364	
Db	1366	cttatct 1372	
RESULT 3			
AAAX56955			
ID	AAAX56955	standard; cDNA: 1260 BP.	
XX	AAAX56955;		
AC			
XX	16-JUL-1999	(first entry)	
DT			
XX			
DE	Human NIP45 coding region.		
XX			
KW	NIP45; human; transcription factor; IL-4; interleukin-4;		
KW	NF-AT interacting protein-45; modulator; treatment; inflammation;		
KW	autoimmune disease; HTP screening; drug testing; allergic; therapy;		
KW	T-cell dependent response; cytokine; diagnostic; immunosuppressant; ss		
XX			
OS	Homo sapiens.		
XX			
PN	W09921993-A1.		
PD	06-MAY-1999.		
XX			
PF	21-OCT-1998; 98WO-GB03141.		
XX			
PR	24-OCT-1997; 97GB-0022388.		
XX			
PA	(ZENNE) ZENNECA LTD.		
PI	Liu D, Zhao J, Zhou H;		
DR	WPI: 1999-312964/26.		
DR	P-PSDB: AAY08330.		
XX			
PT	Polynucleotides encoding human NF-AT interacting protein (NIP45)		
XX			
PS	Claim 2; Fig 2; 84bp; English.		
XX			
CC	This invention describes a novel human NF-AT interacting protein-45		
CC	NIP-45, human NIP45 is a transcriptional trans-activation factor of the		
CC	interleukin-4 (IL-4) gene. Modulators of human NIP45 and IL-4 can be us		
CC	to treat conditions mediated by NIP45 or IL-4, e.g. inflammation or		
CC	autoimmune disease. NIP45 is a good candidate target for HTP screening		
CC	and/or testing system for drugs that will alleviate T-cell dependent		
CC	autoimmune and allergic responses, and for cytokine-based therapies of		

chronic disease. Expression of NRP45 can be inhibited, and IL-4 expression modulated, by administering an antisense molecule. Antibodies against NRP45 and primers can be used in diagnostic assays. Discovery of a human NRP45 is advantageous in that it provides the ability to control IL-4 transcription, which is of importance for anti-inflammation and immunosuppressant drug development.

Sequence 1260 BP; 269 A; 364 C; 407 G; 220 T; 0 other;

Query Match	37.18;	Score 721.6;	DB 20;	Length 1260;
Best Local Similarity	76.48;	Pred. No. 2.3e-192;		
Matches 934; Conservative	0;	Mismatches 264;	Indels 24;	Gaps 3;

OY	48	CCGCGGTGGCCGAGGCGCTCGGAGAGGCCAGGGCCGCTGGCGGTCTCTCGCGCCG	107
Db	45	cggctgcgggcggaggggctggggcgctggggcgctggcgcgctcttcgcggccca	104
OY	108	GCAGCTCTCGGCTAGGCTCATTTCCAGACCGCTGTTGTGACTTGGCTAGCTGACGGA	167
Db	105	gcggtctccatcccgggcagcgctggaagtagtctctgtgacttgttcaaccgcagcga	164
OY	168	CGAAGAGTCTTTGGAAGT-----CGCAGACCCAGTACAGGTGCCGT	205
Db	165	tgaagaaattctgtagtgcgcacccgctcgcgtgtgcgcgcgcgaaggtctgagtgcgc	224
OY	210	GCGCCGCTCCCGCGCGGGGTAAACCTGACACGAGACGCGACAGTACACTGAAAGGGC	265
Db	225	ccggagcccccgggcgcgctgcgctcccggaataacgcacaacgvtacacgcgaaggga	284
OY	270	GGCCGAGGGCGCTGCGGAGGCCCGCGCTACATTGGTGGAGGGCGGCGCGCGCT	325
Db	285	ggaacagcgcgccgaggaaccgcccgcggaagcgctgagcgcgcgcgctgtgtgt	344
OY	330	GGATCCCGGAGAGCGCGCGGTGTGCCAGTGTACTCCGGGAAAGTACAGACACCTCAA	385
Db	345	ggatccggggagggcgcgcgctgtctcgggtgactcggggaagglttaaagcagcttcg	404
OY	390	CCTCATTTCCAGATATATTCCTCCCTTTGAAACTGTGCCCTTCAAGCCTCGAAGTGAAGC	449
Db	405	ccctatcccgatgaltcatctccctcgtgaactctacccctcagggagatgaggaagcgc	464
OY	450	AGATGTGACAAATTCGTGCGAGTTCCTCCTCTGAGAGATGATGCCCTTCAGTTCTCC	509
Db	465	aaagctctgcgaatctc---gagtggtctctacacagagggctcccatcaccacagctctcc	521
OY	510	CTGGGAAAGAGAGCTCAGAAAGAGAGTGTGAAAGAAAGAAATGGAAAGATTTC	565
Db	522	ctggaaagacaagcttgagactaaag---gataaagaaagaaagaaagacagagttctc	578
OY	570	GGACGACGACATCTCTCCTTTGCCCAACCTTGTCTAAGAAAGAAAGCAGAAAGCATAC	625
Db	579	ggatctggaacaactctctctgtccccaacttcaacgaagcacaagcagaacgatac	638
OY	630	GGAGCGCTCCAGAAAGCTAAAGGAAGTGAACAAACGCTCCAAATCTCCGCTCTGCT	688
Db	639	tcgggactccaagaaglttaaigtgaggtgaacaagcgctccaagatctccgttctgtct	699
OY	690	GAGCGCCAGAGACGACGAGTCCAGGCCCTTCAAGAGACAGATGATGAGTGGTCTAGT	749
Db	699	gagccccaagcaccctcaggttcaagagcaacagggccaagaggaatgaagtgtgtctgt	758
OY	750	GGAAGGGCTGTCTTGGCCACAGAGCTCTCGACTTTTACACTCAAGATCCGGGTCCGGC	809
Db	759	ggaagggcccaacctcccaagagaccccccgactcttcccaatcaaaatccgttgcggc	818
OY	810	TGACTATGTGAGACTGCGCTCGAGAGATGTGAGACCCCTTCAGAAATGTGGGTGCAT	869
Db	819	tgaactgttcaatctgcctcccaagatgttcggagccccctgcagaagtgtgtgtgaccaat	878
OY	870	GGCCATCATCTTGGGGTGTCTCCAACAGGATTTCTTTGCTTTTGGAGAGAGTGAAT	929
Db	879	ggccaaccaacttgggtgtctcccaagagaaatctcttgccttcttgadagaaagagct	938

PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Lamson G, Drmanac R, Ckvenjajlov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
DR WPI: 2000-126369/11.
XX
XX Polynucleotide library used to determine cancerous states of mammalian
PT cells -
PS Claim 1: Page 1090; 1097pp; English.
XX
XX AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
CC libraries constructed from human colon cancer cell lines. The present
CC invention also describes a method of detecting differentially expressed
CC genes correlated with a cancerous state of a mammalian cell, comprising
CC detecting at least one differentially expressed gene product in a test
CC sample derived from a cell suspected of being cancerous, where detection
CC of the differentially expressed gene product is correlated with a
CC cancerous state of the cell from which the test sample was derived.
CC The polynucleotides sequences can be used in a method for detecting
CC differentially expressed genes correlated with a cancerous state of a
CC mammalian cell. The polynucleotides can also be used as probes for
CC detecting and mapping related genes. They can be used in diagnosis and
CC prognosis of diseases and disorders (e.g. identification of
CC pre-metastatic or metastatic cancerous states, stages of cancer, or
CC responsiveness of cancer to therapy). This is particularly for breast
CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
CC negative breast cancer, lung cancer, and colon cancer.
XX
XX Sequence 300 BP; 62 A; 98 C; 79 G; 61 T; 0 other;
SQ

Query Match 10.2%; Score 197.6; DB 21; Length 300;
Best Local Similarity 81.0%; Pred. No. 1.1e-45;
Matches 230; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 679 CGCTCTGCTGAGCCCAAGACAGAGTCAGAGCCCTTGCAGAGACAGATGAG 738
II IIIII IIIII IIIII IIIII III III III III III III IIIII
DB 17 cgttcctgtctgagcccaagcactcaggggtcgaagagaacagggccaaagagatgaa 76
QY 739 GTGTCTCTAGTGAAGAGGCTGTCTTCCACAGAGCTCTTGACTTTTACATCAAGATC 798
IIIIII IIIII IIIII III III III III III III IIIII III
DB 77 gtgtcttgtgtgaaagggccaccctccagagagaccccccagactctccaccacaatc 136
QY 799 CGGTGCGGAGCTGACCTAGTGAAGAGTCAGAGATGTCGAGAGCCCTTGCAGATGTG 858
II IIIII IIIII III III III III IIIII IIIII IIIII IIIII
DB 137 cgttgcgaggtctgaactggtcagatgtccctcaggaatgtcggagccctgcagagtg 196
QY 859 GTGTATCATGTGGCCATCTCTTGGGGTGTCTCAAAAGATTTCTTTTGTGTTTGA 918
IIII IIIII IIIII III IIIII IIIII III IIIII IIIII IIIII
DB 197 gtgaccacatgagccacccttggtgggtgtcccaagcagatcccttgccttttga 256
QY 919 GAAGTGAATGTCTCTACTGACACCCCTAGTACCTTAAGCT 962
IIII III III III IIIII IIIII III IIIII IIIII
DB 257 gagacagagctacactactactcactccagagaccctaaagct 300

RESULT 6
AAZ79987/c
ID AAZ79987 standard; cDNA; 705 BP.
XX
XX AAZ79987;
AC
XX
XX 07-APR-2000 (first entry)
DT
XX
XX Human colon cancer cell line SW480 cDNA clone SEQ ID NO:71.
DE
XX
XX Human; gene expression product; diagnosis; tumour; colon cancer;
KW colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;

KW hyperplasia; ds.
XX
XX Homo sapiens.
OS
XX MO9964576-A2.
PN
XX 16-DEC-1999.
PD
XX
XX 09-JUN-1999; 99MO-IB01062.
PF
XX 10-JUN-1998; 98US-0088801.
PR
XX
XX (FARB) BAYER CORP.
PA
XX
XX Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI Carroll E, Catino TJ, Dertl A, Ford DM, Lewis ME, Monahan JE;
PI Schlegel R;
XX
XX WPI: 2000-087220/07.
DR
XX
XX Novel nucleic acids, used to develop products for the diagnosis and
PT treatment of disorders involving unwanted cell proliferation,
PT particularly cancers, especially colon cancer -
XX
XX Claim 1: Page 160; 469pp; English.
PS
XX
XX AAZ79917 to AAZ80766 represent double stranded cDNA clones isolated from
CC the human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC cDNA clones can be used to generate antisense oligonucleotides which
CC can be used for antisense therapy. Methods and products from the present
CC invention can be used for identifying and/or classifying cancerous cells
CC present in a human tumour, particularly in solid tumours, e.g.
CC carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones
CC can be used for developing agents for the diagnosis and treatment of
CC disorders involving unwanted cell proliferation, such as neoplasia,
CC dysplasia or hyperplasia.
XX
XX Sequence 705 BP; 190 A; 134 C; 194 G; 157 T; 30 other;
SQ

Query Match 8.6%; Score 167.6; DB 21; Length 705;
Best Local Similarity 86.4%; Pred. No. 4.7e-37;
Matches 185; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 779 GACTCTTACATCAAGATCCGGTCCGGGCTGACCTAGTGAAGACTGCTGCAGGATGT 838
I IIIII IIIII IIIII IIIII IIIII IIIII III III III IIIII
DB 214 GCCTCTTCCACATCAAAATCCGTTGCGGGCTGACTGTGCAGATGGCCCTCAGGATGT 155
QY 839 CGAGAGCCCTTCAGATGTGTGTGATCAGATGACCAATCATCTTGGGGTGTCTCCAAACA 898
IIIIII IIIII IIIII IIIII IIIII IIIII III IIIII IIIII III
DB 154 CGAGAGCCCTTCGACAGTGTGTGTGACCAATGACCACTTGGGGTGTCTCCAAACA 95
QY 899 GGAATCTTTTGTCTTTTGGAGAGAGTGAAGTGTCTCTACTGACACCCCTAGTACCTTAA 958
IIII IIIII IIIII IIIII IIIII IIIII III IIIII IIIII IIIII
DB 94 GGAATCTTTTGTCTTTTGGAGAGAGACATGATGACCTTGTGCTTGTGCTTGTGCT 35
QY 959 AGCTTGAGTGGCTGACATCATGATGATGTGTGT 992
IIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
DB 34 AGCTCGAGTGGCTGACATCATGATGATGTGTGT 1

RESULT 7
AAQ73500
ID AAQ73500 standard; DNA; 8438 BP.
XX
XX AAQ73500;
AC
XX
XX 15-MAY-1995 (first entry)
DT
XX
XX DNA encoding Pseudorabies virus large latency transcript.
DE
XX
XX Pseudorabies virus; PRV; LRV; large latency transcript;
KW attenuated virus; vaccine; early protein 0; EPO; HSV-1 ICP0;

XX	Key	Location/Qualifiers
FH	misc_feature	1..1078
FT		/*tag= a
FT		/label= St B12.3
FT	misc_feature	1079..1985
FT		/*tag= b
FT		/label= St B12.5
XX		
PN	W09214840-A.	
XX		
PD	03-SEP-1992.	
XX		
PE	13-FEB-1992;	92WO-FR00145.
XX		
PR	13-FEB-1991;	91FR-0001684.
XX		
PA	(HEID/) HEIDET S.	
PA	(INRM) INSERM INST NAT SANTE & RECH MED.	
XX		
PI	Heitz D, Mandel J, Oberle I, Rousseau F, Vincent A;	
PI	Mandel JL;	
XX		
DR	WPI; 1992-316192/38.	
XX		
PT	Nucleic acid fragment of X chromosome region - for diagnosis of	
PT	fragile X syndrome and related mental retardation	
XX		
PS	Claim 8; Fig 8; 53pp; French.	
XX		
CC	Probe 9B12.3 (or "St B12.3") is used to detect anomalies at the	
CC	fragile X site after digestion of genomic DNA by EcoRI or HindIII.	
CC	This and other probes, including St B12.5, which can be used in the	
CC	diagnosis of Fragile X syndrome, were obtained by subcloning DNA	
CC	fragments from the 225kb region between probes St677 and D033. This	
CC	region is localised to the Xq27-q28 region of the human X chromosome	
CC	cloned as an artificial yeast chromosome. Specifically, the probes	
CC	are adjacent to or comprise CpG sites.	
CC	See also AAQ28291-3 and AAQ28295.	
XX		
SQ	Sequence 2183 BP; 520 A; 655 C; 533 G; 460 T; 15 other;	
Query Match 2.1%; Score 41.6; DB 13; Length 2183;		
Best Local Similarity 67.0%; Pred. No. 0.19;		
Matches 59; Conservative 0; Mismatches 29; Indels 0; Gaps 0;		
OY	243 GGACGCGCAGCTGACAGTGAAGAGGGGCGCCAGAGGGCCTGGCGAGAGCCCGCTACATT 302	
Db	1750 GGGCGCGGGCGGTGAGAGGAGCGCGCTCCAGGGGGCGTGGCGAGACCGCGGCGCGCG 1691	
OY	303 GGTGCGACGCGGCGGCGGCGGCTGCTG 330	
Db	1690 GCGCGCGCGCGCGCGGCGGAGGCGGCGG 1663	
RESULT 13		
ID	AAAX23299/C	
XX	AAAX23299 standard; DNA; 3656 BP.	
XX		
AC	AAAX23299;	
XX		
DF	11-JUN-1999 (first entry)	
XX		
DE	Human SPA-1 cDNA.	
XX		
KW	SPA-1; tumour-suppressing; low molecular weight G protein activation;	
KW	treatment; gene therapy; cancer; human; ds.	
XX		
OS	Homo sapiens.	
XX		
key	'	Location/Qualifiers
FH		297..3425
FT	CDS	

FT		/tag= "a"
ET		/product= "SPA-1"
XX		
PN		W09910380-A1.
XX		
PD		04-MAR-1999.
XX		
PF		21-AUG-1998; 98WO-JP03715.
XX		
PR		22-AUG-1997; 97JP-0240218.
XX		
PA		(MINA/) MINATO N.
XX		
PI		Minato N;
XX		
DR		WPI, 1999-190588/16.
XX		P-Psdb; AAW3489.
PT		New human SPA-1 protein - has a strong tumour suppressing capacity
XX		
PS		Claim 1; Page 29-37; 78pp; Japanese.
XX		
CC		This invention describes a novel human SPA-1 protein with a strong
CC		tumour-suppressing activity and which is capable of activating low
CC		molecular weight G proteins. Human SPA-1 is useful for the treatment
CC		(including gene therapy) of cancers in humans.
XX		
SQ		Sequence 3656 BP; 667 A; 1257 C; 1161 G; 571 T; 0 other;
Query Match	2.1%;	Score 40.2; DB 20; Length 3656;
Best Local Similarity	49.4%;	Pred. No. 0.63; Mismatches 133; Indels 4; Gaps 1;
Matches 134; Conservative	0;	
OY	31	GACGTCGGTCCAGGTCTCCGCCGCGTGCGAGGCCTCGAGAGCCCGCAGCCCGTGGC 90
DB	2538	GGAGCTGGGCAACGGGCTCTGGGCGGAGGCTTGGCAGCGTCTGGCGCACAGCGCAAGA 2479
OY	91	CGGTCTCTCGCGCGCCGCGCAGTCTCCGGCTAGGCTCATTTCCAGACAC---CGTCTTGT 146
DB	2478	GCGCGCGCCCGGGCGGCGCAGCCCCCGCGCTCTCGCGGAATGTGAAGCGCTCCAGTGTCTGA 2419
OY	147	GGACTTGGTCAATGACAGCAGAGAGAGTCTTGGAAATCCGAGACCAGTAAGAGTGGC 206
DB	2418	CGAATTCCTCTGGCGGTTCACACTCGAAAGCCAGCGGCTTTGACCGTGGCGGGCAGCGCCA 2359
OY	207	GGTCGCGCCGCTCCCGCGCGGCGCTAAACCTGACGAGACAGCAGTACAGTAGAAGG 266
DB	2358	GCTCGCGGCTCTCGAGACGACGCGCTCACCAGTGTGAGGCGCGCCACCACTTGCCCAAGG 2299
OY	267	GGCGGCGGAGGGGCTTGGCGGAGAGCCCGGT 297
DB	2298	CATTGGCGGGGAGCCGCTCGAAGCGCAGCGT 2268
RESULT 14		
AAT72781/c		
ID	AAT72781	standard; cDNA: 4425 BP.
XX	AAT72781;	
AC		
DT	23-SEP-1997	(first entry)
XX		
DE		Human interleukin-1 receptor interacting protein cDNA.
XX		
KW		Interleukin-1 receptor interacting protein; IL1I; IL-1; inhibitor;
XX		tumour necrosis factor; ds.
XX		
OS		Homo sapiens.
XX		
FH	Key	Location/Qualifiers
FT	CDS	1108..4236
FT		/*tag= a

Wed Sep 12 06:07:09 2001

us-09-192-611-1.rng

Page 12

[illegible]

```
Search completed: September 11, 2001, 12:59:14
Job time: 4501 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 11, 2001, 11:42:58 ; Search time 72.63 Seconds
(without alignments)
5072.277 Million cell updates/sec

Title: US-09-192-611-1
1946
Sequence: 1 ACAGTGTGGAGATGGCGGA.....CGCGATTCGGCGACTCGAG 1946

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA:*
1: /cgn2_6/prodata/2/lna/5A_COMB.seq:*
2: /cgn2_6/prodata/2/lna/5B_COMB.seq:*
3: /cgn2_6/prodata/2/lna/6A_COMB.seq:*
4: /cgn2_6/prodata/2/lna/6B_COMB.seq:*
5: /cgn2_6/prodata/2/lna/PCTUS_COMB.seq:*
6: /cgn2_6/prodata/2/lna/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1946	100.0	1946	2	US-08-755-584-1
2	1946	100.0	1946	3	US-09-192-611-1
3	46.4	2.4	8438	1	US-07-943-283-1
4	44	2.3	1026	4	US-07-751-891B-24
5	44	2.3	4188	4	US-07-751-891B-2
6	44	2.3	4242	4	US-07-705-490-2
7	44	2.3	5222	4	US-07-751-891B-23
8	41.6	2.1	4362	2	US-08-455-073A-1
9	38.6	2.0	1002	2	US-08-890-980-5
10	38.6	2.0	1002	3	US-08-890-979-5
11	38.6	2.0	1002	3	US-09-032-894-5
12	38.6	2.0	1002	3	US-09-032-894-5
13	38.6	2.0	1002	4	US-09-031-626-5
14	38.6	2.0	1002	4	US-09-031-626-95
15	38.6	2.0	246240	2	US-08-724-394A-20
16	38.6	2.0	246240	2	US-08-724-394A-21
17	38.6	2.0	246240	2	US-08-724-394A-22
18	38.4	2.0	7218	1	US-08-232-463-14
19	38.2	2.0	4897	6	5196516-7
20	37.4	1.9	43280	2	US-08-804-227C-1
21	37	1.9	11219	1	US-07-642-734C-1
22	37	1.9	11219	3	US-08-439-009A-1
23	36.2	1.9	2965	3	US-08-906-360-2
24	36	1.8	2219	3	US-08-510-646B-17
25	36	1.8	38506	3	US-09-320-878-19
26	35.8	1.8	44377	2	US-08-804-227C-7
27	35.8	1.8	44377	2	US-08-804-198-1

C	28	35.6	1.8	1931	2	US-09-130-114-2	Sequence 2, Appl
C	29	35.4	1.8	13987	2	US-08-804-227C-13	Sequence 13, Appl
C	30	35.2	1.8	774	3	US-08-956-307B-12	Sequence 12, Appl
C	31	35.2	1.8	778	3	US-08-956-307B-11	Sequence 11, Appl
C	32	35.2	1.8	1788	3	US-08-303-861-1	Sequence 1, Appl
C	33	34.8	1.8	423	1	US-08-470-179-113	Sequence 113, App
C	34	34.8	1.8	1020	2	US-08-475-844-10	Sequence 10, Appl
C	35	34.8	1.8	1020	5	PCT-US95-08429-10	Sequence 10, Appl
C	36	34.8	1.8	8056	3	US-09-136-605-14	Sequence 14, Appl
C	37	34.8	1.8	8082	1	US-08-306-691B-41	Sequence 41, Appl
C	38	34.8	1.8	8082	1	US-08-187-785-1	Sequence 1, Appl
C	39	34.8	1.8	8082	5	PCT-US93-06251-28	Sequence 28, Appl
C	40	34.6	1.8	2091	3	US-08-899-437-22	Sequence 22, Appl
C	41	34.6	1.8	2091	4	US-09-126-121-22	Sequence 22, Appl
C	42	34.6	1.8	2502	3	US-08-899-437-5	Sequence 5, Appl
C	43	34.6	1.8	2502	4	US-09-126-121-5	Sequence 5, Appl
C	44	34.2	1.8	2562	1	US-08-146-421-4	Sequence 4, Appl
C	45	34.2	1.8	12001	1	US-08-458-568A-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-08-755-584-1
; Sequence 1, Application US/08755584
; Patent No. 5858711
; GENERAL INFORMATION:
; APPLICANT: Glimcher, Laurie H.
; APPLICANT: Hodge, Martin R.
; TITLE OF INVENTION: NF-AT-INTERACTING PROTEIN NIP45 AND METHODS
; TITLE OF INVENTION: OF USE THEREFOR
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/755,584
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Giulio A. DeConti, Jr.
; REGISTRATION NUMBER: 31,503
; REFERENCE/DOCKET NUMBER: HUI-026
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1946 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 13..1248
; US-08-755-584-1
Query Match 100.0%; Score 1946; DB 2; Length 1946;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 1946; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 ACAGTGTGGAGATGTGGGAAACCACTGAGGGGACGTGTGCTCCGAGGTCCCGGTGGCCGA 60
    |||||||
Db 1 ACAGTGTGGAGATGTGGGAAACCACTGAGGGGACGTGTGCTCCGAGGTCCCGGTGGCCGA 60
OY 61 GGGGCTCGAGAGACCCGAGGCGCCGCTGGCGGCTGTCTGGCGCCGCGAGTGTCCGGCT 120
    |||||||
Db 61 GGGGCTCGAGAGACCCGAGGCGCCGCTGGCGGCTGTCTGGCGCCGCGAGTGTCCGGCT 120
OY 121 AGGCTCATTTCCAGACACCCGTGTGTGTGAGCTTGTGATGACAGACGACGAAGAGTCTTG 180
    |||||||
Db 121 AGGCTCATTTCCAGACACCCGTGTGTGTGAGCTTGTGATGACAGACGACGAAGAGTCTTG 180
OY 181 GAAGTGGCAGAACCCAGTAGAGGTGCGGCTGCCCGCTCCCGCGCCGCTAAACTGAG 240
    |||||||
Db 181 GAAGTGGCAGAACCCAGTAGAGGTGCGGCTGCCCGCTCCCGCGCCGCTAAACTGAG 240
OY 241 CAGGACAGCGACAGTGCAGTGAAGGGGCGCGAGGGGCGCTCGGGAGGCGCCGCTGACA 300
    |||||||
Db 241 CAGGACAGCGACAGTGCAGTGAAGGGGCGCGAGGGGCGCTCGGGAGGCGCCGCTGACA 300
OY 301 TTGGTGGACGCGCGGCGCGCGCTGTGTGATCCGGAGAGGCGCGGTGTCCAGTG 360
    |||||||
Db 301 TTGGTGGACGCGCGGCGCGCGCTGTGTGATCCGGAGAGGCGCGGTGTCCAGTG 360
OY 361 TACTCGGGAGGTAGACAGCAGCTCAACCTATTCCAGATTAATTCATCCCTTGAAA 420
    |||||||
Db 361 TACTCGGGAGGTAGACAGCAGCTCAACCTATTCCAGATTAATTCATCCCTTGAAA 420
OY 421 CTGTGCCCTTCAGAGCCTGAAGATGAGGAGCATCTGCAAAATTCCTGGCAGTTCCTCT 480
    |||||||
Db 421 CTGTGCCCTTCAGAGCCTGAAGATGAGGAGCATCTGCAAAATTCCTGGCAGTTCCTCT 480
OY 481 GAGAGATGATGCTGCTTGTAGGTCTCCCTGAGAAAGAGTCCAGAAAGAGTGTGAG 540
    |||||||
Db 481 GAGAGATGATGCTGCTTGTAGGTCTCCCTGAGAAAGAGTCCAGAAAGAGTGTGAG 540
OY 541 AAAGAAGAAAGAAATGGAAGATTTCCGGACAGAGCATCTCTCTTTGCCCAACT 600
    |||||||
Db 541 AAAGAAGAAAGAAATGGAAGATTTCCGGACAGAGCATCTCTCTTTGCCCAACT 600
OY 601 TCGTCAAGAAAGAAAGCAGAAAGCATACGAGGCGCTCCAGAAAGTAAAGGAAGTGAAC 660
    |||||||
Db 601 TCGTCAAGAAAGAAAGCAGAAAGCATACGAGGCGCTCCAGAAAGTAAAGGAAGTGAAC 660
OY 661 AACGCTTCCTCAAGATCTCCGCTCTGCTGAGCTCCCAAGAGCAGCAGATCCAGCCCTT 720
    |||||||
Db 661 AACGCTTCCTCAAGATCTCCGCTCTGCTGAGCTCCCAAGAGCAGCAGATCCAGCCCTT 720
OY 721 CAGAGCAGAGATGATGAGTGTCTGTAGTGAAGGGGCTCTTGTGCCACAGAGCTCTGA 780
    |||||||
Db 721 CAGAGCAGAGATGATGAGTGTCTGTAGTGAAGGGGCTCTTGTGCCACAGAGCTCTGA 780
OY 781 CTCTTTACACTCAAGATCCGGTCCGGGCTGACCTAGTGAAGTGCCTGTCAAGATGTG 840
    |||||||
Db 781 CTCTTTACACTCAAGATCCGGTCCGGGCTGACCTAGTGAAGTGCCTGTCAAGATGTG 840
OY 841 GAGCCCTTCAGAAATGTGTGTGATACATGTGCGCAATCATCTTGGGGTGTCTCAAAACAG 900
    |||||||
Db 841 GAGCCCTTCAGAAATGTGTGTGATACATGTGCGCAATCATCTTGGGGTGTCTCAAAACAG 900
OY 901 ATTCTTTTGGCTTTTGGAGAGAGTGAAGTGTCTCTACTGCAACCCCTAGTACCTTAAAG 960
    |||||||
Db 901 ATTCTTTTGGCTTTTGGAGAGAGTGAAGTGTCTCTACTGCAACCCCTAGTACCTTAAAG 960
OY 961 CTGTGAGTGGCTGACATCATTTGATGTGTGTGCTAGCAAGCTTTTCAGAGGCGCACAGAG 1020
    |||||||
Db 961 CTGTGAGTGGCTGACATCATTTGATGTGTGTGCTAGCAAGCTTTTCAGAGGCGCACAGAG 1020
OY 1021 ACATCCAGAGAGTCCGGGCTCCGGGTGCAAGGGGAAGAAACCAACGATGTTGGAGATC 1080
    |||||||
Db 1021 ACATCCAGAGAGTCCGGGCTCCGGGTGCAAGGGGAAGAAACCAACGATGTTGGAGATC 1080
```

```
Db 1021 ACATCCAGAGAGTCCGGGCTCCGGGTGCAAGGGGAAGAAACCAACGATGTTGGAGATC 1080
OY 1081 TCACGTCTCCCTGATCTCTCTTAAGCTTCATGTCACACTATGAGGAAGGCATGGGA 1140
    |||||||
Db 1081 TCACGTCTCTTAATCTCTCTTAAGGTCTCATGTCCACACTATGAGGAAGGCATGGGA 1140
OY 1141 CTCTGTGACACAAAGCTCTCTCTCTTGTGATGGGACAAAGCTTTCAGGCAAGAGCTG 1200
    |||||||
Db 1141 CTCTGTGACACAAAGCTCTCTCTCTTGTGATGGGACAAAGCTTTCAGGCAAGAGCTG 1200
OY 1201 CCAGCTGATCTGGGCTCGAATCCGAGATCTCATGGAAGTGTGGGCTGAAAGCTCTAC 1260
    |||||||
Db 1201 CCAGCTGATCTGGGCTCGAATCCGAGATCTCATGGAAGTGTGGGCTGAAAGCTCTAC 1260
OY 1261 CCGTGTGAGGAGGAAAGCCAGACATGAGACAAATGCTCCCAATTTTATTTGATTT 1320
    |||||||
Db 1261 CCGTGTGAGGAGGAAAGCCAGACATGAGACAAATGCTCCCAATTTTATTTGATTT 1320
OY 1321 TTTTGGCCCATTAAGGGCTAAACAGAACTGAATTAAGAACTTGTACTTATTTATTTCTGG 1380
    |||||||
Db 1321 TTTTGGCCCATTAAGGGCTAAACAGAACTGAATTAAGAACTTGTACTTATTTATTTCTGG 1380
OY 1381 TCGTGGGAGATTGAACCCAGACTATGACATGCTTAAGGATGTATGAAGTGGAGCAAAAC 1440
    |||||||
Db 1381 TCGTGGGAGATTGAACCCAGACTATGACATGCTTAAGGATGTATGAAGTGGAGCAAAAC 1440
OY 1441 CAAGGCAATTAATCTTTAGCCAGCCTAGTAGACTGTAGTCAAGCAAGTGGCTACTTGG 1500
    |||||||
Db 1441 CAAGGCAATTAATCTTTAGCCAGCCTAGTAGACTGTAGTCAAGCAAGTGGCTACTTGG 1500
OY 1501 TAGTGTGTGCTCTGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1560
    |||||||
Db 1501 TAGTGTGTGCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1560
OY 1561 GGAACCTTGGCTTCCCTACCTTTACCTTTACCTTTACCTTTACCTTTACCTTTACCTTT 1620
    |||||||
Db 1561 GGAACCTTGGCTTCCCTACCTTTACCTTTACCTTTACCTTTACCTTTACCTTTACCTTT 1620
OY 1621 CTGTGAAGCTGCTATGTGAGTGTGTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1680
    |||||||
Db 1621 CTGTGAAGCTGCTATGTGAGTGTGTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1680
OY 1681 CTTGGGGTTTGTCCGCAAGTGTGAGCAGCAGCAGTCTGCTGACTGTGGACTGAGAA 1740
    |||||||
Db 1681 CTTGGGGTTTGTCCGCAAGTGTGAGCAGCAGCAGTCTGCTGACTGTGGACTGAGAA 1740
OY 1741 TGCATTTCTCGTGTGAGACACTCGGGTGCAGAAATTAACAGAAAGTGAACATGCTG 1800
    |||||||
Db 1741 TGCATTTCTCGTGTGAGACACTCGGGTGCAGAAATTAACAGAAAGTGAACATGCTG 1800
OY 1801 AACCTGAGCACTAGTGCAGAAAGTTAAGCAAGTGTGCAATTTTACGCTTGGGTATCTCTGT 1860
    |||||||
Db 1801 AACCTGAGCACTAGTGCAGAAAGTTAAGCAAGTGTGCAATTTTACGCTTGGGTATCTCTGT 1860
OY 1861 GCCTGCCAGAGCTCTAGCCAGTGTGTGTACACACTTCTTGGCATGAGACCTAGTGTGA 1920
    |||||||
Db 1861 GCCTGCCAGAGCTCTAGCCAGTGTGTGTACACACTTCTTGGCATGAGACCTAGTGTGA 1920
OY 1921 CCGGGGCGGATTCGGCCGACCTGAG 1946
    |||||||
Db 1921 CCGGGGCGGATTCGGCCGACCTGAG 1946
```

RESULT 2

```
US-09-192-611-1
; Sequence 1, Application US/09192611
; Patent No. 6090561
; GENERAL INFORMATION:
; APPLICANT: Glimecher, Laurie H.
; APPLICANT: Hodges, Martin R.
; TITLE OF INVENTION: NF-AT-INTERACTING PROTEIN NIP4S AND METHODS
; OF USE THEREFOR
; NUMBER OF SEQUENCES: 2
```



```
QY 1561 GGACCTTGGCTCCCTACACATTTACAGTTCGTGTCGTCCTTCTTCATCAGATGACTT 1620
    |||||||
Db 1561 GGACCTTGGCTCCCTACACATTTACAGTTCGTGTCGTCCTTCTTCATCAGATGACTT 1620
QY 1621 CTGTGAAGCTGCTATGTTGAGTGTGTGAACATAAGACCTCTGCTTGGTGTCCAGG 1680
    |||||||
Db 1621 CTGTGAAGCTGCTATGTTGAGTGTGTGAACATAAGACCTCTGCTTGGTGTCCAGG 1680
QY 1681 CCGGGGTTTGTCCCGAGTTTGAGCCAGAGTACCTCTGACTTGGAGTGGAGTGA 1740
    |||||||
Db 1681 CCGGGGTTTGTCCCGAGTTTGAGCCAGAGTACCTCTGACTTGGAGTGGAGTGA 1740
QY 1741 TGCATTTCCCTGTGAGACACTCGGGTGCAGAAATATAACAGAGGATGACATGCTG 1800
    |||||||
Db 1741 TGCATTTCCCTGTGAGACACTCGGGTGCAGAAATATAACAGAGGATGACATGCTG 1800
QY 1801 AAGCTGAGACTAGTTCGAAAATTAAAGACGTTGACTTTTCAGCCTTGGGTATCTCTCT 1860
    |||||||
Db 1801 AAGCTGAGACTAGTTCGAAAATTAAAGACGTTGACTTTTCAGCCTTGGGTATCTCTCT 1860
QY 1861 GCCTGCAGAGCTAGTCCAGTGTCTGTACACACTTGGATGGAGACCTAGTCCGA 1920
    |||||||
Db 1861 GCCTGCAGAGCTAGTCCAGTGTCTGTACACACTTGGATGGAGACCTAGTCCGA 1920
QY 1921 CGCGGGCGCGATTGCGCCGACTCGAG 1946
    |||||||
Db 1921 CGCGGGCGCGATTGCGCCGACTCGAG 1946
```

RESULT 3
US-07-945-283-1
; Sequence 1, Application US/07945283
; Patent No. 5352596

GENERAL INFORMATION:

APPLICANT: Cheung, Andrew K.
APPLICANT: Wesley, Ronald D.
TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
TITLE OF INVENTION: Involving The EP0 and LIT Genes
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSER: Curtis P. Ribando
STREET: 1815 No. 5352596th University Street
CITY: Peoria
STATE: IL
COUNTRY: USA
ZIP: 61604

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/945,283
FILING DATE: 19920911
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Ribando, Curtis P.
REGISTRATION NUMBER: 27976
TELEPHONE: 309-685-4011 ext. 513
TELEFAX: 309-685-4128
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8438 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
*ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Pseudorabies virus

```
FEATURE:
NAME/KEY: CDS
LOCATION: 622..6495
FEATURE:
NAME/KEY: variation
LOCATION: replace(1099, "g")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1267, "t")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1381, "c")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1566, "c")
FEATURE:
NAME/KEY: variation
LOCATION: replace(7010, "g")
```

Query Match 2.4% Score 46.4; DB 1: Length 8438;
Best Local Similarity 49.3%; Pred. No. 0.0049;
Matches 149; Conservative 0; Mismatches 151; Indels 2; Gaps 1;

```
QY 29 GGGAGCTGTGTCCGAGTCCCGCGGTGCGCGAGCGCTCGAGAGCCCGAGCGCCCTG 88
    |||||
Db 5462 GAGAGAGAGGTGCGGAGCGCGAGGCGCGAGGCGCGCGCGCGCGCGCGCGCGCG 5521
QY 89 GCGCGTGTCTCGCGCGCGCGCACTCCGCTAGCTCATTCACAGACACGTCCTGTG 148
    |||||
Db 5522 GAGAGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5581
QY 149 ACTTGTCAGTACAGCAGCAGAGAGTCTTGAAGTCCGACACCCAGTAGAGTCCCG 208
    |||||
Db 5582 TCTCCCTC--GAGAGAGAGCAGAGAGAGAGAGCAGAGAGAGAGAGAGAGAGAG 5639
QY 209 TCGCCCGCTCCCGCGCGCGCTAAACCTGAGCAGACAGCAGACTGACAGTGAAGG 268
    |||||
Db 5640 CCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5699
QY 269 CGCGCGAGGGGCGCTCGCGAGAGCCCGCTGATTTGTCGACCGCGCGCGCGCTG 328
    |||||
Db 5700 CGCGGAGAGTCCGAGTCCGCTCGCGCGCTGCTGAGCTGTGAAGCGCGCGCGCG 5759
QY 329 TG 330
    |
Db 5760 GG 5761
```

RESULT 4

US-07-751-891B-24
; Sequence 24, Application US/07751891B
; Patent No. 6180337

GENERAL INFORMATION:

APPLICANT: Caskey, C. T.
APPLICANT: Nelson, David L.
APPLICANT: Pieretti, Maura
APPLICANT: Warren, Stephen T.
APPLICANT: Oostera, Ben A.
Fu, Ying-hui
TITLE OF INVENTION: Diagnosis of the Fragile X Syndrome
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSER: Thomas D. Paul
STREET: 1301 McKinney, Suite 5100
CITY: Houston
STATE: Texas
COUNTRY: U.S.A.
ZIP: 77010-3095

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

LENGTH: 4242 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOHETICAL: NO
US-07-705-490-2

Query Match
Best Local Similarity 59.7%; Score 44; DB 3; Length 4242;
Matches 74; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 243 GGACAGCGACGTGACAGTGAAGGGGGCGGAGGGGCTGCGGAGAGCCCGCGTACATT 302
DB 1642 GGGCGCGCGCGGTGACGAGAGCGCGCGCTGCGAGGGGCGTGGGAGAGCGGCGCGCGC 1701
QY 303 GGTGCGACGGCGGCGGCGGCGGCTGCTGATCCCGAGAGAGCGCGCGGTGCTCCAGTGA 362
DB 1702 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTGCGGCTGAGAGCGCGCGAGCC 1761
QY 363 CTCG 366
DB 1762 CACC 1765

RESULT 7
US-07-751-891B-23
Sequence 23, Application US/07751891B
Patent No. 6180337
GENERAL INFORMATION:
APPLICANT: Caskey, C. T.
Nelson, David L.
Pieretti, Maura
Warren, Stephen T.
Oostera, Ben A.
Fu, Ying-hui
TITLE OF INVENTION: Diagnosis of the Fragile X Syndrome
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Thomas D. Paul
STREET: 1301 McKinney, Suite 5100
CITY: Houston
STATE: Texas
COUNTRY: U.S.A.
ZIP: 77010-3095
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/751,891B
FILING DATE: 29-Aug-1991
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Paul, Thomas D.
REGISTRATION NUMBER: 32,714
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713/651-5325
TELEFAX: 713/651-5246
TELEX: 762829
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 3222 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-07-751-891B-23

Query Match
Best Local Similarity 59.7%; Score 44; DB 4; Length 5222;
Matches 74; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 243 GGACAGCGACGTGACAGTGAAGGGGCGGAGGGGCTGCGGAGAGCCCGCGTACATT 302
DB 2671 GGGCGCGCGGTGACGAGAGCGCGCGCTGCGAGGGGCGTGGGAGAGCGGCGCGCGC 2730
QY 303 GGTGCGACGGCGGCGGCGGCGGCTGCTGATCCCGAGAGAGCGCGCGGTGCTCCAGTGA 362
DB 2731 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTGAGAGCGCGCGAGCC 2790
QY 363 CTCG 366
DB 2791 CACC 2794

RESULT 8
US-08-455-073A-1
Sequence 1, Application US/08455073A
Patent No. 5876949
GENERAL INFORMATION:
APPLICANT: Gideon Dreyfuss
APPLICANT: Mikiko C. Stomi
APPLICANT: Yan Zhang
TITLE OF INVENTION: Fragile X Related Proteins, Compositions And Methods
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5876949r1s
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/455,073A
FILING DATE: 31-MAY-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Deluca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: UPN-2201
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4362 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 220..2118
US-08-455-073A-1

Query Match
Best Local Similarity 67.0%; Score 41.6; DB 2; Length 4362;
Matches 59; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 243 GGACAGCGACGTGACAGTGAAGGGGCGGAGGGGCTGCGGAGAGCCCGCGTACATT 302
DB 12 GGGCGCGCGGTGACGAGAGCGCGCGCTGCGAGGGGCGTGGCGCAAGCGGCGGCGCGC 71

TITLE OF INVENTION: Sequences and Antibodies thereo
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 336
ATTORNEY/AGENT INFORMATION:
NAME: Filts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0300
TELEFAX: 415-576-0200
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1..246240
OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-20

Query Match 2.0%; Score 38.6; DB 2; Length 246240;
Best Local Similarity 54.1%; Pred. No. 4.5;
Matches 100; Conservative 0; Mismatches 84; Indels 1; Gaps 1;
QY 1041 CCGGGTCAGAGGAGGAGAAACAGATGTTGAGATCTCAGCTGCTCCTGATTCTCC 1100
Db 113699 CAGAGTTTGGGACAGAGAAACCATCTTGGAGTCTTAAGTCTTAAGAGGTGTC 113758
QY 1101 TCTTAAGTTCTCATGTCACACTATGAGAGAGCCATGGAAGTCTCTGACACAGACTCTC 1160
Db 113759 CCTTAGACATCTAGTCCGCTGTTATGAGGTCTGCTGCTGAGAGAGAGCCCTT 113818
QY 1161 CTCTCTTGTGATGGAGCAAGCTTTCAGGCAAGAGAGCTGCCAGTATCTGAGGCTCGA 1220
Db 113819 GTTGATGATGAG-GAGCAAAATTTTAAACAAGAACTGTACCCGTATCTGGGCATGAA 113877
QY 1221 ATCCG 1225
Db 113878 GCCAG 113882

Search completed: September 11, 2001, 12:56:56
Job time: 4438 sec

This Page Blank (top)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 11, 2001, 08:48:26 : Search time 17.87 seconds
(without alignments)
1756.235 Million cell updates/sec

Title: US-09-192-611-2

Perfect score: 2099
Sequence: 1 MAEPLRGPRGRGRRGARR.....GKELPADLGESGLIEVWG 412

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	154.5	7.4	1217	2 T00270	hypothetical prote
2	145.5	6.9	1072	1 A37221	neurofilament trip
3	140	6.7	1634	2 T26517	hypothetical prote
4	137.5	6.6	854	2 S02003	neurofilament trip
5	134.5	6.4	587	1 WMBETE	65k early nonstruc
6	133	6.3	971	2 T19431	hypothetical prote
7	132.5	6.3	792	2 T29187	hypothetical prote
8	132	6.3	586	1 WMBEDE	65k early nonstruc
9	128.5	6.1	635	2 T09648	nucleolin homolog
10	128.5	6.1	734	2 B42680	nucleolin homolog
11	128	6.1	303	1 Z8BE22	suppressor two of
12	127	6.0	1365	2 S14871	hypothetical prote
13	126	6.0	856	2 T16543	hypothetical prote
14	124.5	5.9	1684	2 JMO057	gravin - human
15	124.5	5.9	1791	2 T02345	hypothetical prote
16	122	5.8	1082	2 T15269	hypothetical prote
17	120	5.7	669	2 JC5662	hepatoma-derived g
18	119	5.7	450	2 T08701	hypothetical prote
19	119	5.7	765	2 T49592	neurofilament trip
20	119	5.7	1603	2 S17983	gene posterior sex
21	119	5.7	1611	2 T34836	hypothetical prote
22	118.5	5.6	508	2 T04605	hypothetical prote
23	118	5.6	577	2 T12536	hypothetical prote
24	117.5	5.6	1110	2 I51116	NF-180 - sea lamp
25	117.5	5.6	2187	2 T30826	nascent polypeptid
26	117	5.6	553	2 C75318	hypothetical prote
27	116.5	5.6	714	2 T22454	hypothetical prote
28	116.5	5.6	955	2 S52959	male-specific leth
29	116.5	5.6	974	2 A40580	lodestar maternal-

30	116	5.5	672	2 H86169	hypothetical prote
31	116	5.5	1043	2 A56037	DNA-binding protei
32	115.5	5.5	872	2 S62061	SCD5 protein - yea
33	115.5	5.5	1280	2 T00365	hypothetical prote
34	115	5.5	382	2 T14336	RAD23 protein, iso
35	114	5.4	891	2 G84693	probable prolina-r
36	114	5.4	1274	2 T16251	hypothetical prote
37	113.5	5.4	2361	2 T25752	hypothetical prote
38	113	5.4	611	2 T06458	nucleolin homolog
39	113	5.4	776	2 T20738	hypothetical prote
40	112.5	5.4	95	2 JC4760	SMI3 protein - hum
41	112.5	5.4	675	2 T03744	myod protein inhib
42	112.5	5.4	2109	2 T33247	hypothetical prote
43	112.5	5.4	3942	2 T42730	Bassoon protein -
44	112	5.3	480	3 JC7552	Shb-like adapter p
45	112	5.3	900	2 C96842	hypothetical prote

ALIGNMENTS

RESULT 1

T00270
hypothetical protein KIA0596 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C:Accession: T00270
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, DNA Res. 5, 31-39, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The comp
A:Reference number: Z14086; MIMID:98290545
A:Accession: T00270
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 11217 <NAG>
A:Cross-references: EMBL:AB011168; NID:93043715; PIDN:BAA25522.1; PID:93043716
A:Experimental source: brain
C:Genetics:
A>Note: KIA0596

Query Match 7.4%; Score 154.5; DB 2; Length 1217;
Best Local Similarity 23.5%; Pred. No. 0.086;
Matches 105; Conservative 48; Mismatches 164; Indels 129; Gaps 21;

QY	6	RGRGPRSRGGRGARRARGARGRCPRAROSPRLIPDVIVDIVSDSEVLEVADPVEVP	65
DB	455	RGRGKQGGPSSPQRASG-----PNRHQAPSMISPGPA-----LSDSDEGEDETEELP	506
QY	66	-----VARLPAPAKPEQSDSDSEGAEE-----GPGAPRTLVRRRRRL	106
DB	507	ALPYLAKSTKKALASVPSPALPRSLSHWMSRAQESGFIDPAPANAPG-----RRRRRW	563
QY	107	DPGEPAPVVPYSGKVGSSLLIPDNSSLKLCPS--EP-EDPADLITNSGSSPSDDALPS	163
DB	564	QPG-----VELSVRSMIDLRLQLETLAPSLAQPSODSLAIFSGRRKHQGLALET	612
QY	164	G-----SPWRKKLRKKC-----EKPE-----KMEPEPPDDISPLPOPS--SRNK	201
DB	613	SLTSQNEKPPRPOASQPCSTPHIIRLLSQEBGVFAQULEPAPIEDGIVYEPPSPNPMDT	672
QY	202	SRKHTALQK-----LREVNKRLOD-----LRSCLSPKQKOSPALQSTDEEVL	245
DB	673	SEFQVQAPARAGTLGRVYPGSRSSSEKHPDSCSVSYSSCLSSPEH--PTEDSESTPLS	730
QY	246	VEGPVLPOSSRLFTLKIRCADLVRLPYRMSEPLQNVVDMAHNLGVSPNRILLFESE	305
DB	731	VDG-----ISSDLEEPGEDEEEEGEGMGF-----YGLQF	762
QY	306	LSPATP-----STLKIGVADITDCVVLASSEATETSOELRLRYOGKKEHMLE	355
DB	763	GSP-QTPDQEQFLKQHETTLASCAAPGAP--VQYPERSESISIRFLLOYOTRLREP--	818

R:Breen, K.C.; Robinson, P.A.; Wion, D.; Anderton, B.H.
 FEBS Lett. 241, 213-218, 1988
 A:Title: Partial sequence of the rat heavy neurofilament polypeptide (NF-H). Identificat
 A:Reference number: S02003; MUID:89065087
 A:Accession: S02003
 A:Molecule type: mRNA
 A:Residues: 1-854 <BREF>
 A:Cross-references: EMBL:X13804; NID:957828; PIDN:CAA32038.1; PID:957829
 C:Superfamily: neurofilament triplet H protein
 C:Keywords: coiled coil

Query Match 6.6%; Score 137.5; DB 2; Length 854;
 Best Local Similarity 21.9%; Pred. No. 0.6;
 Matches 94; Conservative 62; Mismatches 168; Indels 105; Gaps 20;

```

QY 4 PLRGRGRSGRGARARAGRCRPARARL-IPDYLVLVSDSEEV----- 55
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 372 PVEAKSPAERKSPASVSPG-EAKSPAERKSPAEVKSPTVKSPEAKSPAEVKSPTVK 430
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 56 -LEVADPVEVPVAPRLPAPAKPEODSDSEGAEGPAGARTLVRRRRRLDPGEAPV 113
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 431 SPRAKSPVEY-----KSPASVKSPEAKSPAGAKSPAEKSPVYAKSPAEKSPAGAK- 484
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 114 VPVYSGVQSSLNLPNLSILKLPSEPEDEADLTN-----SGSSPSEDA--LP 162
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 485 -PRAEAKSPAERKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEK 538
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 163 SGSPWRKKLRKCKEKKMEFPDODISPL-PPSSKNSR-----KHTELQKLRVKN 217
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 539 AKSEPKASPVKEIKPAPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEK 598
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 218 RLQDLRCLSPKQSPALQSTDEEVLVVEGVLPOSRLFTLKIRCADLVRLP--VR 274
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 599 RPADIR---SPEQVKSRAKEAKS-----PEKEFTTEKAVAKKEVKSPEVEVK 645
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 275 MSEPLQNVDMANHLGVSPRILLFGESELSPTAPSTLKLGVADIIDCVLASSSEA 334
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 646 AKEP-----PKV-----EEEKTP-ATPKTE-----VKEKSKDEA 674
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 335 TETSQELRLRVQGEKQHMLEISLSPDSPLVKMSHYEAMGLSGHKLSFFDGTKL 394
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 675 PKEAQ---KPAEKEKPELVE--KPKDSPGEAKKEAEKKAAP-----E 715
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 395 ELPADLGL 403
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 716 ETPAKLGVK 724
| : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 5

WMBETE

65K early nonstructural protein - human cytomegalovirus (strain Towne)
 C:Species: human cytomegalovirus, human herpesvirus 5
 A:Note: host Homo sapiens (man)
 C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Jul-1999
 C:Accession: A41808
 R:He, Y.S.; Xu, L.; Huang, E.S.
 J. Virol. 66, 1098-1108, 1992
 A:Title: Characterization of human cytomegalovirus UL84 early gene and identification of
 A:Reference number: A41808; MUID:92114132
 A:Molecule type: mRNA
 A:Residues: 1-587 <HEV>
 A:Cross-references: GB:M81332; NID:g330541; PIDN:AAA45947.1; PID:g330542
 C:Superfamily: human cytomegalovirus 65K early nonstructural protein
 C:Keywords: leucine zipper; nonstructural protein
 F:114-135/Region: leucine zipper motif
 F:325-373/Region: leucine zipper motif

Query Match 6.4%; Score 134.5; DB 1; Length 587;
 Best Local Similarity 24.4%; Pred. No. 0.58;
 Matches 105; Conservative 58; Mismatches 159; Indels 109; Gaps 22;

```

QY 6 RGRGRSR-----GGRGARARAGRCRPARAR-----PARLIPTVVLVLD--VSDSEEV 55
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 11 RARPRARAGCGGCGVSGNSNRHS--CKRRQRALAPLFTLATTTTMMGVASTDDSD 69
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 56 LEVADPVEVPVAPRLPAPAKPEODSDSEGAEGPAGARTLVRRRRRLDPGEAPV 115
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 70 LLKTPPEL-----DKHSGSPQITL-----TLDKHDHIDR 101
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 116 VYSGK---VQSSLNLPN-----SSLKLCPSPEDEADLTNGSSPSDDALPS 163
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 102 VHRGTYLILQHLDRPELDPQIILLSTPLQGEAGESQ---TAPATQEEETA-- 156
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 164 GSPRRKLRKCKEKKMEFPDODISPLPPSSKNSRKHTALQKLRVKNRLQDLR 223
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 157 -----SHELEKKKEKKEKEE--DED-----DRNDRR-----GILCVSMDSDVR 198
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 224 SCLG--PKQOSPALQSTDEEVLVVEGVLPOSRLFTLKIRCADLVRLVRSSEPLQN 281
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 199 PAFSLFPAKPCGCHILRSYIDQ-QLRMAIVRLSNLFLRL-ITPPLKRVPLRRAAHHT 256
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 282 VV-DHMANHLGVSPNRIILFGESELSPTAT-PSYTKL-GVADIIDCVLASSSEATETS 338
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 257 ALHDCMALHL-----PELTFTSLDINNVTENAAVADAESTDADLT 299
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 339 QELRLRVQGEKQHMLEISLSPDSPLVK-----LMSHYEAMGLSGHKLSFFDGTKL 393
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 300 PTLVTVRVNHAVCWHRVEGCGISGPRGLTSRISARLSEYTKAKLGPS-----VFGRL 353
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 394 KELPADLGL 404
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 354 NESPDILTSS 364
| : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 6

T19431

hypothetical protein C25A1.10 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
 C:Accession: T19431
 R:Mortimore, B.
 submitted to the EMBL Data Library, October 1996
 A:Reference number: Z19124
 A:Accession: T19431
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-971 <WLB>
 A:Cross-references: EMBL:Z81038; PIDN:CAE02755.1; GSPDB:GN00019; CESP:C25A1.10
 A:Experimental source: clone C25A1
 C:Genetics:
 A:Gene: CESP:C25A1.10
 A:Map position: 1
 A:Insertions: 38/3; 92/3; 201/3; 919/3
 C:Superfamily: nucleolus-cytoplasm shuttle phosphoprotein

Query Match 6.3%; Score 133; DB 2; Length 971;
 Best Local Similarity 23.3%; Pred. No. 1.3;
 Matches 60; Conservative 35; Mismatches 115; Indels 48; Gaps 8;

```

QY 29 PRARQSPARLIPDYLVLDVS---DSDEVELEVADPVEVPVAPRLPAPAKPEODSDSE 85
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 231 PAAPKTPAKPTPKPVYKAESSSDSDSEKPKVAKP--APAKATPKPAKKAADSSDSD 288
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 86 -----AAGSPGAPRTLVRRRRRLD-----PEEAVVPVYSGKVOSSLNLI 128
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 289 DEAPAKKTPAKAAPVAKKAESSSDSDSEKPAKPTPKAKATPKVAKKAESSSDSD 348
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 129 PDNSLILKLPSE--PEDEADLTNGSSPSDDALPSGSPW--RKRLRCKEKKMEFP 185
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 349 DEKPPVAKPAPAKATPKVAKKAESSSDSDSEKPAKPTPKAKATPKVAKKAESSSD 408
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 186 PDODISPLPPSSRNKS-----RKHTALQKLRVKNRLQDLRSC 225
| : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db 409 SDDKKKVVAPRTSAKATPKPAKKRADSSSDDEAPAKKTPAKAAKPAKKAEE---SS 465
 OY 226 LSPKQHSPALQSTDDDEV 243
 Db 466 SDSSDDEKPAKSTPAKI 483

RESULT 7

T29187

hypothetical protein C55C3.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T29187

R:Moessene, J.; Steillyes, L.

Submitted to the EMBL Data Library, April 1996

A:Description: The sequence of C. elegans cosmid C55C3.

A:Reference number: 220585

A:Accession: T29187

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-792 <MOE>

A:Cross-references: EMBL:U53335; PIDN:AAA96170.1; GSPDB:GN00022; CESP:C55C3.3

A:Experimental source: strain Bristol N2; clone C55C3

C:Genetics:

A:Gene: CESP:C55C3.3

A:Map position: 4

A:Introns: 17/1; 46/3; 139/2; 173/1; 204/3; 250/3; 266/3; 286/2; 327/1; 362/3; 393/3; 43

Query Match 6.3%; Score 132.5; DB 2; Length 792;
 Best Local Similarity 20.9%; Pred. No. 1.1;
 Matches 90; Conservative 50; Mismatches 135; Indels 155; Gaps 19;

OY 11 RSRGGRGARRARAGRCPRARQSPARLIPDTVLVDLSDEEVLEVPVVARLP 70
 Db 309 RRRGGNOIKKKPKPK---PLKRITP-----NNSDEKLKKKSP----- 343
 OY 71 APAPKPEDSDSD-----SEGAAGPAGAPRTLVRRRRR-----LLDPG 109
 Db 344 -PKKPLQATOSDILQPLPSNNAMTVPESSKESANNSPOSSKINLVNRNGRVTNTDLDCS 402
 OY 110 EAPVVPYVSGVQSSSLMLIPNSSLKLCPSPEDEADLTNGSSPEEDALPGSGPWK 169
 Db 403 EARAIP-----KC-DKLDFQFTSTNSEQSPA--DAISAVAP-TK 437
 OY 170 KLRKCKEKKKMEFPD-----ODI-----SPLPOSSRNK----- 201
 Db 438 LTGQSQSEKQEIEMKNSFRELODSKFKPAEAFIPKPEERKGSTQSTTEPTKLGSQ 497
 OY 202 SRKHTALQKLRVYNKRLQD--LRSCLSPKQHOSPALQSTDD----- 241
 Db 498 SPSEKQEIQMPNFSFRELODSKFKPAPSEKSPALQSESPFVPGNNRRNSETSOVD 557
 OY 242 -----EVVLVEGPVLPQSSRLFTLKIRCADLVRLPKRKSEPLQVNVDH-----MAN 288
 Db 558 TISPVPTKLVGTQSPSEKQEIEMPNFSFRELODSKFKPAPKPKPEVERGSAEQSMSE 617
 OY 289 HLGSPRRILLFGESSELPATPSTLK-----LGVAIDIICVLA-----SS 331
 Db 618 PLGTS-----RYARG-SPIAKRPRSPQLQAPLETLATPTTIDAPTAIETAIENSAEFSS 672
 OY 332 SEATETSQEL 341
 Db 673 SHSMDPNSNL 682

RESULT 8

WM8DE

65k early nonstructural protein - human cytomegalovirus (strain AD169)

N:Alternate names: UL84 protein

C:Species: human cytomegalovirus, human herpesvirus 5

A:Note: host Homo sapiens (man)

C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Jul-1999
 C:Accession: S09848
 C:Cheer, M.S.; Bankier, A.T.; Beck, S.; Bohm, R.; Brown, C.M.; Cerny, R.; Horsnell, T
 M.; Barrell, B.G.
 Curr. Top. Microbiol. Immunol. 154, 125-169, 1990
 A:Title: Analysis of the protein-coding content of the sequence of human cytomegalovi
 A:Reference number: S09749; MUID:90269039
 A:Accession: S09848
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-586 <CHE>
 A:Cross-references: EMBL:X17403; NID:959591; PIDN:CAA35358.1; PID:91780862
 A:Note: this sequence was submitted to the EMBL Data Library, December 1989
 C:Superfamily: human cytomegalovirus 65k early nonstructural protein
 C:Keywords: leucine zipper; nonstructural protein
 F:114-135/Region: leucine zipper motif
 F:324-372/Region: leucine zipper motif

Query Match 6.3%; Score 132; DB 1; Length 586;
 Best Local Similarity 23.6%; Pred. No. 0.82;
 Matches 104; Conservative 62; Mismatches 145; Indels 130; Gaps 25;

OY 6 RGRGPRSR---GGGARRARAGRCPRARQSPARLIPDTVLVDL--VSDSEV 55
 Db 11 RARRRARRRGGGGGGGSNSRHS-GKCRORRALSAPLFTLTATTTTMMGVASTDDDS 69
 OY 56 LEVADPVEVVARLPAPAKPEQSDSDSEGAABGAPAGAPRTIVRRRRRLLDGEAVVP 115
 Db 70 LLKTPDEL-----DKYSGSPOTIL-----TLDKHDIRQPR 101
 OY 116 VYSK---VQSSINLIPDN-----SSLLKLCPSPEDEADLTNGSSPEEDALPS 163
 Db 102 VHRGTYHLQLHDLRPEELRDPQILSTPQLQ--GEANDESG--TAPATLDEEETAS 157
 OY 164 GSPWRKRLKCKEKKKMEFPDQDISPLPQSSRNKSKKHTDALQKREYNKRLQDLR 223
 Db 158 HEPEKKK---EKQEKKEE---DED-----DRNDDKER---GLICVSNSESDVR 197
 OY 224 SCLS--PKQHSRALQSTDEEVLVGEPVLPQSSRLFTLKIRCADLVRLPKRKSEPLQ 281
 Db 198 PAFSLFPARGCHILKRVYDQ-QLTRMAIVRLSINLFALPI-ITPLKRLPLR-----RK 250
 OY 282 VYDHMANHLGVSPNRILLFGESSELPATPSTLKLGVAIDIICVLAASSEATETSQEL 341
 Db 251 AANHSTALH-----DCLALHLPDLTFEPTLDINNTEAASVAD---TAESTDA-DLTPTL 301
 OY 342 RLRYQCKEKHOMLEISLSPSPKLVLMSHYEAMGLSG-----HKLS----- 383
 Db 302 TVRVR-----HALCWHVEG-GISGPRGLTSRISARLSETTAKTLGP 342
 OY 384 FFDGTYLKGKELPADLGLES 404
 Db 343 SVFGRLDELDPNESPDLTLSS 363

RESULT 9

T09648

nucleolin homolog num1 - alfalfa

N:Alternate names: num1 protein

C:Species: Medicago sativa (alfalfa)

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000

C:Accession: T09648

R:Boegre, L.; Jonak, C.; Mink, M.; Meskienne, I.; Traas, J.; Ha, D.T.C.; Swoboda, I.;

Plant Cell 8, 417-428, 1996

A:Title: Developmental and cell cycle regulation of Alfalfa nucleolin a plant homolog of

A:Reference number: Z16796; MUID:9631876

A:Accession: T09648

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-635 <BOE>

A:Cross-references: EMBL:X08845; NID:q1279562; PIDN:CAA61298.1; PID:q1279563

C:Genetics:

A;Gene: num1
C;Superfamily: nucleolin; ribonucleoprotein repeat homology

Query Match	6.1%	Score 128.5;	DB 2,	Length 635;
Best Local Similarity	20.8%	Pred. NO. 1.5;		
Matches 92; Conservative	75;	Mismatches 183;	Indels 93;	Gaps 20

```

Oy 24 ARGCGPARQSPA-----RLIDPVLVDVSD---SDEEVLVADPVVPAARLPAPAK 74
      ! : : : : : ! : : : : : ! : : : : : ! : : : : : ! : : : : :
Db 80 AKAPAP-SKTRPANNGVYKKAQPTTISESDSDDISDEEYKPPAKAVPSKNSGAPVK 138
      ! : : : : : ! : : : : : ! : : : : : ! : : : : : ! : : : : :
Oy 75 -----PEODSDSDSEGAEBGPACAPRTLVRRRRRLLDPEAGVVPVYSGKVOSSLINTLP 129
      ! : : : : ! : : : : ! : : : : ! : : : : ! : : : : ! : : : :
Db 139 KDDESSSESDSESSSESDERKKPAAKAVPSKN-----GSAPAKKDSESEDESDSE 191
      ! : : : : ! : : : : ! : : : : ! : : : : ! : : : : ! : : : :
Oy 130 DNSSLIKLPSE-----PEDEADLUNSGSPSEDDALPS-----GSPMKRKLKKEC 176
      ! : : : : ! : : : : ! : : : : ! : : : : ! : : : : ! : : : :
Db 192 DEKPAKAAVSSKNSGVSAAKKDDSDSDSDSESSSDEEKKPAKASKNSVSAPTKKAASSDE 251
      ! : : : : ! : : : : ! : : : : ! : : : : ! : : : : ! : : : :
Oy 177 KEERKMEEPDODISPLQPS--RNSRKHTTEALOKLREYNKRLQDLRSCLSPKHQSPA 235
      ! : : : : ! : : : : ! : : : : ! : : : : ! : : : : ! : : : :
Db 252 ESDESESD--DEDAKPVSKPAAVAKSKKSDSDDEDDSDSDDEDDKPPAVAKKEVSESE 309
      ! : : : : ! : : : : ! : : : : ! : : : : ! : : : : ! : : : :
Oy 236 LQSTDDDEVVLVEGVLPQSSRLFTLTKIRCADLVRLPYMSEPLQNVVD-----HMANHL 290
      ! : : : : ! : : : : ! : : : : ! : : : : ! : : : : ! : : : :
Db 310 SDSSDDDDHKM--NIDKSDSDSDSESESEDEDELTPOKKIDV--MDVAGSGKAPNT 365
      ! : : : : ! : : : : ! : : : : ! : : : : ! : : : : ! : : : :
Oy 291 GVSPENRILLTGESELSPTATPSTLK-----GVADIICVVLASSSE----- 333
      ! : : : : ! : : : : ! : : : : ! : : : : ! : : : : ! : : : :
Db 366 PATNN-----EISSGSKTLFVGNLSIDYKRSDIENFPQGCYEVVD-VLLADGCGYKRG 417
      ! : : : : ! : : : : ! : : : : ! : : : : ! : : : : ! : : : :
Oy 334 -----ATETSQELRLRVQKKE-KHOMLEIST-----SPDSPLKVLMSHYEAMGLSG 379
      ! : : : : ! : : : : ! : : : : ! : : : : ! : : : : ! : : : :
Db 418 FGHYEFPATAEAQAQALTEMMNQELLHRLRLDLAREGATFPNN-----NSNYSAOSGGRG 472
      ! : : : : ! : : : : ! : : : : ! : : : : ! : : : : ! : : : :
Oy 380 HKLSFFPDG--TKLSGKELPADL 400
      ! : : : : ! : : : : ! : : : : ! : : : : ! : : : : ! : : : :
Db 473 QSOITVFWRGDKNLGDEDEIRAKL 495
      ! : : : : ! : : : : ! : : : : ! : : : : ! : : : : ! : : : :

```

RESULT 10

nucleolus-cytoplasm shuttle phosphoprotein - rat

N:Alternate names: Nopp140 protein B

C:Species: *Rattus norvegicus* (Norway rat)

C:Date: 30-Sep-1993 #sequence-revision 30-Sep-1993 #text-change 17-Nov-2000

C:Accession: B42680; A42680; S27890; S27889; S30510; S30511

R:Meier, U.T.; Blobel, G.
Cell 70, 127-138, 1992

A:Title: Nopp140 shuttles on tracks between nucleolus and cytoplasm.

A:Reference number: A42680; MUID:92323342

A:Accession: B42680

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-734 <MEI>

A:Cross-references: GB:M94288

A:Experimental source: clone PTM6

A:Accession: A42680

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 32-180, 'Q', 181-734 <ME3>

A:Cross-references: GB:M94287; NID:9205749; PIDN:AAA41718.1; PID:g205750

R:Meier, U.; Blobel, G.

Submitted to the EMBL Data Library, May 1992

A:Description: Nopp140 shuttles on tracks between nucleolus and cytoplasm.

A:Reference number: S27889

A:Accession: S27890

A:Molecule type: mRNA

A:Residues: 32-734 <ME2>

A:Cross-references: EMBL:M94288; NID:g205751; PIDN:AAA41719.1; PID:g205752

A:Accession: S27889

A:Molecule type: mRNA

A;Residues: 32-180, 'Q',181-734 <MEI2>
A;Cross-references: EMBL:M94287; NID:9205749; PIDN:AAA41718.1; PID:9205750
C;Superfamily: nucleolus-cytoplasm shuttle phosphoprotein
C;Keywords: phosphoprotein

Query Match	6.1%	Score 128.5;	DB 2;	Length 734;
Best Local Similarity	23.2%;	Pred. No. 1.8;		
Matches 71;	Conservative 35;	Mismatches 105;	Indels 95;	Gaps 12;

[illegible]

RESULT 11
28BP22
gene 8 protein - phage p22
C:Species: Phage P22
A:Note: host *Salmonella typhimurium*
C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 23-Jul-1999
C:Accession: D40474
R:Eppler, R.; Wyckoff, E.; Goates, J.; Parr, R.; Casjens, S.
Virology 183, 519-538, 1991
A:Title: Nucleotide sequence of the bacteriophage P22 genes required for DNA packaging
A:Reference number: A40474; MUID:91306435
A:Accession: D40474
A:Molecule type: DNA
A:Residues: 1-303 <EPP>
A:Cross-references: GB:M59749; NID:g215302; PIDN:AAA72962.1; PID:g553007
C:Genetics:
A:gene: 8
C:Superfamily: phage P22 gene 8 protein

	Query Match	6.1%;	Score 128;	DB 1;	Length 303;	
	Best Local Similarity	21.3%;	Pred. No. 0.64;			
	Matches	73;	Conservative	49;	Mismatches 100;	Indels 120; Gaps 15;
Oy	41 DTVAVDVS-	-DEVELEVADPEVVPVAPRLPAKPEODSDSDEGAEGPAGAPRTLV	98			
	: :	: :				
Db	25 DSLVVDNNDNANGOEGEIV-	-----LDDETAPKQDPKAKNEFRRRRIE	69			
Oy	99 RRRRRLDPCEAPVPYPSGKGOSLNLIPNSSLLKCPSEPEDEADLTNGSGSSPED	158				
	: :	: :				
Db	70 RKROELBQMEA	--VKRGLEPESLAVNP------LPQPPIINVLSEEGLAKEYD	118			
Oy	159 P--AL----	PGSPWRKKLR-----KKECKEKMEEPFDDISLPQSSNNKSARKHTE	207			
Db	119 DNSRLAALFNANNTIMLKKAQDARNSNAVEGRKTQEFQSAOXY-----EAARKHYD	172				
Oy	208 ALQIKREVNKRLODIRSLSPKHOSHOPALQSTDTDEVILVEGEFVLPOSSRLFTLKITCRAD	267				
	:	: :				

[illegible]

```

11
Db 551 ASGE 554

RESULT 13
116543
hypothetical protein K03C7.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Mar-2000
C:Accession: J16543
R:Leimbach, D.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid K03C7.
A:Reference number: Z18532
A:Accession: J16543
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-856 <LEF>
A:Cross-references: EMBL:U00059; NID:g1055170; PID:g1055171; PIDN:AA01138.1; CESP:K03C7.1
C:Genetics:
A:Gene: CESP:K03C7.1
A:Introns: 176/1; 632/2; 653/1; 717/3; 746/3; 791/3
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo

Query Match 6.08; Score 126; DB 2; Length 856;
Best Local Similarity 21.48; Pred. No. 3;
Matches 85; Conservative 55; Mismatches 153; Indels 104; Gaps 17;

QY 2 AEPLRGRRPRSGRGARRARGRCPRARQSPARLLIPDIVLVSDSEVLEVADP 61
Db 382 ADPL--GGPKKKDKPLAKKAPAKK---PTDKPKPK-----DAPKDAKPTPEEPKAP 427
QY 62 V-----EVPARLLPAPAKPQDSDSDEGAEGPAGA-----PRTIVRR 100
Db 428 VAPKMKKRPWEDEPDEPEPEADFTMPAKKPRDIEDPAD-----PLGGPNKKDKPLAKKA 482
QY 101 RRRRLD--FGEARVVPVSGKVOSSLNLPDSSS-----LLKLCSPSEDEADLTNSGS 153
Db 483 PTKKPADPKPSEEPKPVAPAKKMPWEEDPDDEADFTVPPIKPGDEDEPEDADEE 542
QY 154 SPSEDDALPSSGPMKKIKLKCKEKKMEKFEPPRODISLPQRP----- 196
Db 543 EPED--PADEDPKKKPKPKHKKRKKKPVPEPEKPEPEPVVPAKPKIADIKKPEE 599
QY 197 -----SSRNKSKRHTAQLQKLEVNKKRLQDLRSCSPKQHSQAL----- 236
Db 600 PIPMPKPKKTIAERKKKEKRIERIPALRYAKK-PRELEVYIPVPIPE-QTALLITQEGCAAF 657
QY 237 --OSTDDEVVLEGVLP--OSSRLFTL-----KIKRADLVRLVPMSEPLQNVVDH 285
Db 658 GKSRAANVEVNFQDKPIVQAGVDSKTVIPLWNDESKANRSGMTAFGA-FREIDQNVVDH 716
QY 286 MANHIGVSPNRIILLFGESELSPTATPSTLKGVADI 322
Db 717 --HGKSGQIIPLLAKGTYHPHGEYGTIRROTADV 750

RESULT 14
JW0057
gravin - human
C:Species: Homo sapiens (man)
C:Date: 13-Jun-1998 #sequence_revision 10-Jul-1998 #text_change 08-Oct-1999
C:Accession: JW0057
R:Sato, N.; Kokame, K.; Shimokado, K.; Kato, H.; Miyata, T.
J. Biochem. 123, 1119-1126, 1998
A:Title: Changes of gene expression by lysophosphatidylcholine in vascular endothelia
A:Reference number: JW0057; MUID:98269042
A:Accession: JW0057
A:Molecule type: mRNA
A:Residues: 1-1684 <SAT>
A:Cross-references: DDBJ:AB003476; NID:g2081606; PIDN:BA01927.1; PID:d1020716; PID:g
Comment: This protein regulates cell growth.
```

F:433-439/Region: nuclear location signal
 F:522-527/Region: nuclear location signal
 F:591-596/Region: nuclear location signal
 F:671-676/Region: nuclear location signal

Query Match 5.98; Score 124.5; DB 2; Length 1684;
 Best Local Similarity 20.7%; Pred. No. 8.5;
 Matches 94; Conservative 70; Mismatches 171; Indels 119; Gaps 21;

QY 43 VLVNDSDEEVLEVPVEVVARLPAPAKPEDSDSD----- 82
 DB 32 VVHDTTDDGGEPETITQIPSSSESNLEETQPTESQANDIGFKVVEKFKFTVKDKD 91
 QY 83 -----SEGAAGPAGA-----PRTLVRRRRLLDPGEPAPVYVYSGK-VQSS 124
 DB 92 TEKPDVTQLLVKKDEGGAAGAGDHKPSLGAAGAAKSEPKOSTKPEETILKREQSH 151
 QY 125 LNLIP--DNSSLKLCPSPEPD--EADLTNCGSPSEDALPSCSP-----WRK 169
 DB 152 AEISPPASGGAVECEKEGEKEKESKSAESPTSPVTSSTGTFKKFTQGWAGWRK 211
 QY 170 K--LRKKKEK-----EKMEEPDODISPLPQSSRNKSKKTEALQKLRVYKRLQDLR 223
 DB 212 KTSFKPKPEDEVEASEKKKEOEPEK---VDTEEDGKAEEVASEKLTASEQAHPQ--EPAE 265
 QY 224 SCLSPK--OHOSPALOSTDEEVLVVEGP---VLPOSSRLFTLKIRCADLVRLPVRMS 276
 DB 266 SAHEPRLSAETKEVELPS--EROVSGSGSPSEKPAPLATEVFDEKIEVHOEEVAAEVHVS 324
 QY 277 EPLQNVVDHMANHLGVSPNRIILLFGESELSPTATPTTLKLGVAIDIICVVIASSSEATE 336
 DB 325 TVEERTEEQ-----KTEVEETA-----GSVAEELVEMDAEPQEA 360
 QY 337 TSQEL-RLR--VQKKEKQMLEISLSPDPL-----KVLMHYEAPAGLSGHL 382
 DB 361 PAKELVTLKETVCVSGEDPTQAGD--LSPDEKVLKPRGVVSEVEMLSOERMKVQSGSL 418
 QY 383 SFFPDGT---KLSGKELPADLG---LESGDLIEV 410
 DB 419 KKLFTSTGLKLTSGKKKGRGGDESEGHYQV 452

RESULT 15
 T02345
 hypothetical protein KIA0324 - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 05-Nov-1999
 C:Accession: T02345
 R:Ricke, D.O.; Bruce, D.; Muntz, M.; Doggett, N.; Munk, C.; Saunders, E.; Robinson, D.;
 re, J.; White, S.; Ueng, S.; Tatum, O.; Campbell, C.; Fawcett, J.; Deaven, L.
 Submitted to the EMBL Data Library, March 1998
 A:Description: Sequencing of human chromosome 16p13.3.
 A:Reference number: Z14564
 A:Accession: T02345
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1791 <RIC>
 A:Cross-references: EMBL:AC004493; NID:g2996648; PTDN:AC08453.1; PTD:g2996650
 C:Genetics:
 A:Map position: 16
 A:Introns: 1610/2; 1706/2
 A:Note: KIA0324

Query Match 5.98; Score 124.5; DB 2; Length 1791;
 Best Local Similarity 25.3%; Pred. No. 9.2;
 Matches 72; Conservative 26; Mismatches 96; Indels 91; Gaps 13;

QY 34 SPARLIPVTIVLVSDDEEVLEVPVEVVARLPAPAKPEDSDSDSEGAAGPAGA 93
 DB 414 SSELSPDAVEKAGSSNQ-----SISFVLDAVPTPTSRER-----SSASSPEKMDGL 463

QY 94 PRTLVRRRRLLDPGEPAPVYVYSGKVQSSLNLIPDNSSL-----LKLCPSEPD 144
 DB 464 PRTPSRRSR-----SGSSPGLRDGSG-----TPSRHSLSGSSPGMKDIPRTPSRGRS 510
 QY 145 EADLTNCGSPSEDALP-----SGSPWRKKLRKKC---EKEKKMEEPDQ----- 188
 DB 511 ECD-----SSP-EPRALPOTPPRPSRSPSPPLNKKCLTPQRENSGSSSVQKTVARTP 564
 QY 189 -----DISPLPQSSR-----NKSRRKTEALQKLRVYKRLQDLRSCLS 227
 DB 565 LQGRSRSSSQEELDYKPSASFOERSESDSPDSKAKTRTPIQRGRSGSSPEVDSKRLS 624
 QY 228 PKHQSS-----PALQSTDEEVLVVEGP---VLPOSSR 256
 DB 625 PRSRSGSSPEYKDKPRAAPRAQSGSDSSPEPKAPAPRALPRSR 669

Search completed: September 11, 2001, 08:49:55
 Job time: 89 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 11, 2001, 08:49:31 ; Search time 10.66 Seconds

(without alignments)
1323.946 million cell updates/sec

Title: US-09-192-611-2

Perfect score: 2099

Sequence: 1 MAEPLRGPRGPRGGRARR.....GKELPADLGLSGDLIEWWG 412

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	135.5	6.5	831	1 NFH_RAT	P16884 ratius norv
2	134.5	6.4	587	1 UL84_HCMVT	P29839 human cytom
3	132	6.3	586	1 UL84_HCMVA	P16737 human cytom
4	128.5	6.1	704	1 NP14_RAT	P41777 ratius norv
5	128	6.1	303	1 VG08_BPP22	P26748 bacterioph
6	127	6.1	1365	1 SUZ2_DROME	P25172 drosophila
7	122.5	5.8	1781	1 AKAC_HUMAN	Q02952 homo sapien
8	121	5.8	1109	1 TCF8_RAT	Q62947 ratius norv
9	119	5.7	1603	1 PSC_DROME	P55820 drosophila
10	116.5	5.6	1039	1 MS11_DROME	P50535 drosophila
11	116.5	5.6	1061	1 LDS_DROME	P34739 drosophila
12	116	5.5	1043	1 TCF8_MESAU	Q60542 mesocricetu
13	115.5	5.5	872	1 SCDS_YEAST	P34728 saccharomyc
14	114.5	5.5	1411	1 TCOF_HUMAN	Q13428 homo sapien
15	113.5	5.4	1280	1 DYN1_RAT	P28023 ratius norv
16	112.5	5.4	95	1 SM32_HUMAN	P55855 homo sapien
17	112	5.3	919	1 DN1L_HUMAN	P18858 homo sapien
18	112	5.3	943	1 CENC_HUMAN	Q03188 homo sapien
19	112	5.3	1004	1 PTPY_RAT	Q63475 ratius norv
20	111.5	5.3	1170	1 XPG_MOUSE	P35689 mus musculu
21	111	5.3	741	1 BSG2_DROME	P33027 zea mays (m
22	111	5.3	897	1 EPI5_MOUSE	P42557 mus musculu
23	111	5.3	1281	1 DYN1_MOUSE	Q08788 mus musculu
24	110.5	5.3	919	1 SYN2_HUMAN	Q9Y618 h nuclea
25	110.5	5.3	2517	1 NCR2_HUMAN	Q9Y618 h nuclea
26	110	5.2	542	1 TUL1_HUMAN	Q00294 homo sapien
27	110	5.2	612	1 ARRS_MAIZE	P13027 zea mays (m
28	110	5.2	883	1 E7AB_DROME	P11536 drosophila
29	110	5.2	1206	1 FM14_MOUSE	Q05839 mus musculu
30	110	5.2	2150	1 SDG3_CAEEL	P34760 caenorhabd
31	109.5	5.2	536	1 GAG_MIVCB	P27460 cas-br-e mu
32	109	5.2	974	1 YMB1_CAEEL	Q03601 caenorhabd
33	109	5.2	1089	1 Y553_HUMAN	Q9ukj3 homo sapien

34	108.5	5.2	487	1 ATF2_CHICK	Q93602 gallus gall
35	108.5	5.2	505	1 TUB_MOUSE	P50566 mus musculu
36	108.5	5.2	1117	1 TCF8_MOUSE	Q64318 mus musculu
37	108	5.1	487	1 ATF2_HUMAN	P15336 homo sapien
38	108	5.1	487	1 ATF2_MOUSE	P16951 mus musculu
39	108	5.1	673	1 FXR2_HUMAN	P51116 homo sapien
40	108	5.1	1402	1 IF4G_RABIT	P41110 oryctolagus
41	107.5	5.1	532	1 ICP0_HSVB	P28990 equine herp
42	107	5.1	487	1 ATF2_RAT	Q00969 ratius norv
43	107	5.1	650	1 NDC1_XENTIA	P20397 xenopus lae
44	107	5.1	1053	1 DYN1_CHICK	P35458 gallus gall
45	106.5	5.1	610	1 ARLC_MAIZE	P13526 zea mays (m

ALIGNMENTS

RESULT	1	STANDARD	PRT	831 AA.
ID	NFH_RAT			
AC	P16884: 063368:			
DT	01-AUG-1990 (Rel. 15, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)			
DE	(NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H) (FRAGMENT).			
GN	NEFH OR NFH.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain:			
RX	MEDLINE=89065087; PubMed=3143606;			
RA	Breen K.C., Robinson P.A., Wion D., Anderson B.H.;			
RT	"Partial sequence of the rat heavy neurofilament polypeptide (NF-H).			
RT	Identification of putative phosphorylation sites.";			
RL	FEBS Lett. 241:213-218(1988).			
RN	[2]			
RP	SEQUENCE OF 37-831 FROM N.A.			
RX	MEDLINE=88309090; PubMed=2457365;			
RA	Deutigny A., Pham-Dinh D., Roussel C., Felix J.M., Nussbaum J.L.,			
RA	Jolles P.;			
RT	"The large neurofilament subunit (NF-H) of the rat: cDNA cloning and			
RT	in situ detection.";			
RL	Biochem. Biophys. Res. Commun. 154:1099-1106(1988).			
RN	[3]			
RP	SEQUENCE OF 1-89 AND 243-313 FROM N.A.			
RX	MEDLINE=87080760; PubMed=2878828;			
RA	Robinson P.A., Wion D., Anderson B.H.;			
RT	"Isolation of a cDNA for the rat heavy neurofilament polypeptide			
RT	(NF-H).";			
RL	FEBS Lett. 209:203-205(1986).			
RN	[4]			
RP	SEQUENCE OF 318-831 FROM N.A.			
RX	MEDLINE=89184647; PubMed=2928342;			
RA	Lieberburg I., Spinner N., Snyder S., Anderson J., Goldhaber D.,			
RA	Smolowitz M., Carroll Z., Emanuel B.S., Breithner J., Rubin L.;			
RT	"Cloning of a cDNA encoding the rat high molecular weight			
RT	neurofilament peptide (NF-H): developmental and tissue expression in			
RT	the rat, and mapping of its human homologue to chromosomes 1 and			
RT	22.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).			
CC	-I- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,			
CC	AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.			
CC	NEF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT			
CC	SUBSERVED BY THE TWO SMALLER NF PROTEINS.			
CC	-I- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPLETIC K-S-P, NFH IS			
CC	PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS			
CC	THOUGHT THAT PHOSPHORYLATION OF NEF RESULTS IN THE FORMATION OF			
CC	INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE			
CC	OF AXONAL CALIBER.			

-1- PM PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND COINCIDENTLY WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.

-1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.

-1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 783 ONWARD AND IS LONGER DUE TO A FRAMESHIFT.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M37227; AAAA1693.1; ALT_FRAME.
DR EMBL; X13804; CAA32038.1; ALT_FRAME.
DR EMBL; M21964; AAAA1695.1; ALT_FRAME.
DR EMBL; J04517; AAAA1692.1; -.
DR PIR; A30796; A30796.
DR PIR; A25649; A25649.
DR PIR; B25649; B25649.
DR PIR; S02003; S02003.
DR InterPro: IPR001664; -.
DR Pfam; PF00038; filament; 1.
DR ProSite; PS00226; IF_1.
KW Intermediate filament; Heptad repeat pattern; Coiled coil; Neurone;
KM Phosphorylation; Repeat.
FT NON TER 1
FT DOMAIN 1
FT CONFLICT 276 641 51 x 3 AA TANDEM REPEATS OF K-S-P.
FT CONFLICT 164 164 L->I (IN REF. 2).
FT CONFLICT 185 185 I->S (IN REF. 2).
FT CONFLICT 193 193 L->T (IN REF. 2).
FT CONFLICT 199 199 M->T (IN REF. 2).
FT CONFLICT 346 346 K->N (IN REF. 1).
FT CONFLICT 373 373 A->V (IN REF. 1 AND 4).
FT CONFLICT 482 482 G->E (IN REF. 2 AND 4).
FT CONFLICT 485 485 P->S (IN REF. 2).
FT CONFLICT 570 571 RK->KE (IN REF. 2 AND 4).
FT CONFLICT 591 591 P->T (IN REF. 2 AND 4).
FT CONFLICT 727 727 A->V (IN REF. 4).
FT CONFLICT 757 759 AAP->GST (IN REF. 4).
FT CONFLICT 757 759 T->L (IN REF. 2).
FT CONFLICT 769 769 R->P (IN REF. 2 AND 4).
FT CONFLICT 775 775 R->P (IN REF. 2 AND 4).
SQ SEQUENCE 831 AA; 89486 MM; 1B0973C3F13EF768 CRC64;

Query Match 6.5%; Score 135.5; DB 1; Length 831;
Best Local Similarity 21.9%, Pred. No. 0.39;
Matches 94; Conservative 61; Mismatches 169; Indels 105; Gaps 20;

QY 4 PLRGGRSGRGRARRARGRCPPRAROSPRL-IPDTIVLVDSDEEV----- 55
Db 372 PAEAKSPAEKASPAVKSPG-EAKSPAEEKSPAEEKSPATVRSPEAKSPAEEKSPVTYK 430
QY 56 --LEVADVEVPVARLPAPAKPEODSDSEGAAGCAPARTIVRRRRRLLDGEAPV 113
Db 431 SPAEKKSVEV-----KSPASYKSSPEAKSPAGAKSPAEEKSPVYAKSPAEEKSPAGAK- 484
QY 114 VPVYSKGVOSSLNLIPDNSSLIKLCPSPEDEADLTN-----SGSSPSEDDA---LP 162
Db 485 -PPAEKSPAEEKSPAEEKS-----PAEAKSPAEEKSPVEVSPAEKASPVYEKAISLA 538
QY 163 SGSPRRKLRLKKCEEEKMEFPPODLSPL-PQSSRNKSR----KHTEALOKLRVNK 217
Db 539 AKSPEKASPYKEETLKPAEVKSPKASPMKKEAKSPKATTLVDYKSPKAPPAKEAK 598
QY 218 RLQDRLSLCPKROHSPALOSTDDEVVLVEGVLPFOSSRLFLTKTRCRADVLRP---VR 274
Db 599 RPADR-----SPQVYKSPAEEKS-----PEKETREKVAKPKEKYKSPVEEVK 645
QY 275 MSEPLQNVDMANHLVGSPNRLLLFGESELSPATPSTLIKLVADIIDCVLIASSSEA 334

```

Db      646 AKEP-----PKKV-----EEEKTP-ATPKTE-----VKESKDEA 674
QY      335 TETSELRTRVQKKERHOMLETSIPDSPKLVLMNSHYEANGLSOHLKISFFPDGKTLGSK 394
Db      675 PKEAQ-----KPKAEKEPELTE--KPKDSPGPAKKPEAKEKKAAPAE-----E 715
QY      395 ELPAADLGLE 403
Db      716 ETPAKLGVK 724

RESULT 2
ID      UL84_HCMVT STANDARD: PRT: 587 AA.
AC      P29839;
DT      01-APR-1993 (Rel. 25, Created)
DT      01-APR-1993 (Rel. 25, Last sequence update)
DT      01-APR-1993 (Rel. 25, Last annotation update)
DE      65 KDA EARLY NONSTRUCTURAL PROTEIN (UL84 PROTEIN).
GN      UL84.
OS      Human cytomegalovirus (strain Towne).
OC      Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC      Betaherpesvirinae; Cytomegalovirus.
OX      NCBI_TaxID=10363;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=92114132; PubMed=1309892;
RA      He Y.S., Xu L., Huang E.S.;
RT      "Characterization of human cytomegalovirus UL84 early gene and
RL      J. Virol. 66:1098-1108(1992).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; M81432; AAA45947.1; -.
DR      PIR; A41808; WMBETE.
KW      Nonstructural protein.
FT      DOMAIN 9 19 ARG-RICH (BASIC).
FT      DOMAIN 162 170 LYS-RICH (BASIC).
FT      DOMAIN 171 183 ASP/GLU-RICH (ACIDIC).
SQ      SEQUENCE 587 AA; 65388 MW; 13C10E41FB3220B CRC64;

Query Match 6.4%; Score 134.5; DB 1; Length 587;
Best Local Similarity 24.4%; Pred. 0.29;
Matches 105; Conservative 58; Mismatches 159; Indels 109; Gaps 22;

QY      6 RGRGRSR-----GGRGARRAGARGRCPRANOS---PALIPIPTVLVDL--VSDSDEEV 55
Db      11 RARRRRARRGGGGVGSNSSRHS-GCKRRORRALSAPIPLTEIATTTTMMGVASTDDDS 69
QY      56 LEVADPVVPAVARLPAKPKPEDSDSDSGAEGPAGAPRTLVRRRRRLIDPGAPVPV 115
Db      70 LLTKPTDEL-----DKHSGSPOTL-----TLTDKHDHROP 101
QY      116 VYSCK---VQSSLANLIPN-----SLTLKLCPSPEDEADLTGSSGSPSEDALPS 163
Db      102 VHRGTYHLIQHLIDRLPELNDPQIILSTPLQGLGANGESQ---TATATQDEBETA-- 156
QY      164 GSPWKRLLRKKCEKKEEKEEFPDODISLPDPSSRNKSRKHTAOLKLYREYNKRLOLR 223
Db      157 -----SHELEKKEKKEEKEE--DED-----DRNDRER-----GILCVSNEDSDVR 198
QY      224 SCLS--PRQHSAPALQSTDDDEVLYEGSVFLQSSRLFLTKIRCRADIVRLVVRKSEPLQN 281
Db      199 PAFSLFPARPGCHILIRSVIDQ-QLTRMAIVIRLSLMLFLRI--TTPRLKRVPLRKKAAHNT 256

```


FT REPEAT 479 490 ACIDIC SERINE CLUSTER 9.
 FT REPEAT 524 535 ACIDIC SERINE CLUSTER 10.
 FT REPEAT 559 570 ACIDIC SERINE CLUSTER 11.
 FT MOD_RES 567 567 PHOSPHORYLATION (BY CK2).
 FT VARIANT 150 150 MISSING (IN NOBPL40B).
 SQ SEQUENCE 704 AA; 73562 MW; 14DF1BF2DE483EA3 CRC64;

Query Match 6.1%; Score 128.5; DB 1; Length 704;
 Best Local Similarity 23.2%; Pred. No. 0.83;
 Matches 71; Conservative 35; Mismatches 105; Indels 95; Gaps 12;

QY 33 QSPARLIPPTVLVDLSDDSE-----VLEVADPVEVPVRL--PAPAKPEDSDSDSE 84
 DB 317 QSPKKAATOTOPADSSADSSSESESEKTPAKTYVSKTPAKAPAKKKAESSSDS 376
 QY 85 GAAGAGAPRTLVRRRRRLDPGEAPVYV-----YSGKVQSSLNT 127
 DB 377 DSDSDEAPAKPVSAATKSLSPKPAVTPKPPAKAVATPKQAPAGSGQKPSRKADSSSE 436
 QY 128 IPDNS-----LKLCPSPDEA-----DLTNGSSPSSED----- 159
 DB 437 EESSSSSEETATKSVTPPAKVATAKAPSLPAKQAPRAGDSSSDSSSESEKTPPK 496
 QY 160 -----ALPSGSPMKRLRK-----KCEKEKKMEPPDODISPLPQ----- 195
 DB 497 PPAKKAAGAAYPPKPPVKAALAESSSSSSSDSSSESEK-----PKSATPKPAQKA 552
 QY 196 ---PSSRN-KSRKHTPALOKLREVNKRLDLSCLSPKQHSPLASTDDEVVLVGPVL 251
 DB 553 NGVPASQNKAGKESEEDTEONKKAAGTKPGSGKKRKHN----ETADEA-----AT 602
 QY 252 POSSRL 257
 DB 603 POSKKV 608

RESULT 5

VG08_BPP22
 ID VG08_BPP22 STANDARD; PRT; 303 AA.

AC P26748;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE SCAFFOLDING PROTEIN (PROTEIN GP8).
 GN 8.
 OS Bacteriophage P22.
 OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Podoviridae.
 OX NCBI_TaxID=10754;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-15.
 RX MEDLINE-91306435; PubMed-1853558;
 RA Eppeler K., Wyckoff E., Coates J., Parr R., Casjens S.;
 RT "Nucleotide sequence of the bacteriophage P22 genes required for DNA
 packaging.";
 RT Virology 183:519-538(1991).
 RL [2]
 RP SEQUENCE FROM N.A.
 RA Kropinski A.M.B., VanderBijl C.S.;
 RT "The completed sequence of genome of Salmonella phage P22.";
 RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: REQUIRED FOR SUCCESSFUL CONDENSATION OF DNA WITHIN THE
 CAPSID. THE INTERIOR OF THE PROHEAD IS FILLED WITH THE GP8
 PROTEIN. THE SCAFFOLDING PROTEIN IS LOST FROM THE STRUCTURE
 DURING PACKAGING.
 CC -1- INDUCTION: THE SCAFFOLDING PROTEIN NEGATIVELY REGULATES ITS OWN
 SYNTHESIS WHEN IT IS NOT ASSEMBLED INTO PROHEADS.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: M59749; AAA72962.1; -
 DR EMBL: AF217253; AAF75046.1; -
 DR PIR: D40474; Z8BP22.
 KW Late protein.
 SQ SEQUENCE 303 AA; 33564 MW; B86BD625D5011ABE CRC64;

Query Match 6.1%; Score 128; DB 1; Length 303;
 Best Local Similarity 21.3%; Pred. No. 0.32;
 Matches 73; Conservative 49; Mismatches 100; Indels 120; Gaps 15;

QY 41 DYVLVDLSDS--DEEVLVADPVEVPVRLPAPAKPEDSDSDSGAAGPAGARTLV 98
 DB 25 DSLVVDNADNMGQEGFEIV-----LKDDETAPKQDPKKNMEFARRKE 69
 QY 99 RRRRRRLDPGEAPVYVYSGKVQSSLNLIPDSSILKCPSEPEDEADLTNGSSPSSED 158
 DB 70 RKRQRELEQMEA-----VKRGELPESLVNP-----LPQPDINAYLSEGLAKYDY 118
 QY 159 D---AL---PSGSPMKRLK---KCEKEKKMEPPDODISPLPQSSRNKSKKTE 207
 DB 119 DNGRALAFAFNAANTEWLMKAQDARSNAVAEOGRKTOEFTQOSAOYV-----EAARKHYD 172
 QY 208 ALQKLEKVRKRLDLSCLSPKQHSPLASTDDEVVLVGPVLPOSSRLFTLKICRGAD 267
 DB 173 AAEKL-----NIPDYQEKEDAPQVLPPAV-----GAD 200
 QY 268 LVRLPYRMEPLQNVVDHMANHIVGSPNR---ILLFGESELSPTATPSTLKLGVADIID 324
 DB 201 IMRLPEKGAALM-----YHGANPEKARQGLANDGOSAL----- 235
 QY 325 CVVLASSSEATETSOELRLRVQCKEKHQMLETSLSP--DSPL 364
 DB 236 -----IELTRLSERLTLKPRGK-----QISSAPPADQPI 264

RESULT 6

SU22_DROME
 ID SU22_DROME STANDARD; PRT; 1365 AA.

AC P25172;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE SUPPRESSOR 2 OF ZESTE PROTEIN (PROTEIN POSTERIOR SEX COMBS).
 GN SU(2)2.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CANTON-S;
 RX MEDLINE-91279476; PubMed-2057369;
 RA Brunk B.P., Adler P.N.;
 RT "The sequence of the Drosophila regulatory gene Suppressor two of
 RT zeste.";
 RL Nucleic Acids Res. 19:3149-3149(1991).
 CC -1- FUNCTION: REGULATES EXPRESSION OF THE HOMEOTIC SELECTOR GENES BY
 CC INFLUENCING HIGHER-ORDER CHROMATIN STRUCTURE THROUGH INTERACTION
 CC WITH OTHER PROTEINS.

 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 CC -1- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>


```

RESULT 9
PSC_DROME STANDARD; PRT; 1603 AA.
ID PSC_DROME
AC P35820;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE POSTERIOR SEX COMB PROTEIN.
GN PSC.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP MEDLINE=92018190; PubMed=1833647;
RA Brunk B.P., Martin E.C., Sharp E., Adler P.N.;
RT Drosophila genes Posterior Sex Combs and Suppressor two of zeste
RT encode proteins with homology to the murine bml-1 oncogene."
RL Nature 353:351-353(1991).
CC -1- FUNCTION: THE POLYCOMB GROUP (PC-G) GENES ARE NEEDED TO MAINTAIN
CC EXPRESSION PATTERNS OF THE HOMOTIC SELECTOR GENES OF THE
CC ANTENNAPEDIA (ANTP-C) AND BITTORAX (BX-C) COMPLEXES, AND HENCE FOR
CC THE MAINTENANCE OF SEGMENTAL DETERMINATION.
CC -1- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X59275; CAA41965.1; -
DR PIR: S17983; S17983.
DR FLYBase: FBgn0005624; Psc.
DR InterPro: IPR001841; -.
DR Pfam: PF00097; zf-C3HC4; 1.
DR PROSITE: PS00518; ZINC_FINGER_C3HC4; 1.
KW Zinc-finger; Developmental protein; DNA-binding; Nuclear protein.
FT DOMAIN 47 53 POLY-THR.
FT DOMAIN 83 88 POLY-THR.
FT DOMAIN 91 98 POLY-THR.
FT DOMAIN 145 152 POLY-THR.
FT DOMAIN 184 202 POLY-SER.
FT ZN_FING 265 303 POLY-SER.
FT DOMAIN 642 651 POLY-SER.
FT DOMAIN 1066 1069 POLY-SER.
FT DOMAIN 1185 1189 POLY-SER.
FT DOMAIN 1214 1217 POLY-PRO.
FT DOMAIN 1391 1396 POLY-PRO.
FT DOMAIN 1458 1461 POLY-ALA.
FT DOMAIN 1517 1520 POLY-GLY.
SQ SEQUENCE 1603 AA; 169999 MW; 77024f409736473 CRC64;

Query Match 5.78; Score 119; DB 1; Length 1603;
Best Local Similarity 18.9%; Pred. No. 8.3;
Matches 68; Conservative 54; Mismatches 123; Indels 114; Gaps 13;

```

```

DB 715 SPERTLNRAITPPSPVQOASAP-----KSKGNLDDSTLMKPPSCMPKRSIASTRKS 769
QY 203 RKHTALQKLRVKNKRQDLRSLSPKROHSPALQSTDEEVL-----VEG 248
DB 770 KEPKAVASK-----KQKLSPLPTVDFKRLPLVTNGSSGTASPKIEK 812
QY 249 PVLFQSSR---LFTLKTRCADLVRLP-----YMSSE-----LQNVVDHMA- 287
DB 813 PLMPPPAKPMLAPRKLOPSAQFAPPSPIHHAGVOMSAFGNRTPIAKRYOPIIPKASR 872
QY 288 -NHGVSPNRIILLF-----GESELSPTATPST 314
DB 873 PNPEANIPDVNRLLDKAGTEIKSIGGSVYENNSNAQKPHLYGPKGTGMPPALPAT 931

RESULT 10
MSL1_DROME STANDARD; PRT; 1039 AA.
ID MSL1_DROME
AC P50535;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MALE-SPECIFIC LETHAL-1 PROTEIN.
GN MSL-1.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE OF 85-1039 FROM N.A.
RA STRAIN=CANTON-S;
RX MEDLINE=93314941; PubMed=8325488;
RA Palmer M.J., Mergner V.A., Richman R., Manning J.E., Kuroda M.I.,
RA Lucchesi J.C.;
RT "The male-specific lethal-one (msl-1) gene of Drosophila melanogaster
RT encodes a novel protein that associates with the X chromosome in
RT males."
RL Genetics 134:545-557(1993).
RN [2]
RP REVISIONS, SEQUENCE FROM N.A.
RX MEDLINE=95300219; PubMed=7781064;
RA Kelley R.L., Solovyeva I., Lyman L.M., Richman R., Solovyev V.,
RA Kuroda M.I.;
RT "Expression of msl-2 causes assembly of dosage compensation
RT regulators on the X chromosomes and female lethality in Drosophila."
RL Cell 81:867-877(1995).
CC -1- FUNCTION: THE MSL PROTEINS ARE ESSENTIAL FOR ELEVATING
CC TRANSCRIPTION OF THE SINGLE X CHROMOSOME IN THE MALE (X CHROMOSOME
CC DOSAGE COMPENSATION). MSL-1 IS A PIONEER PROTEIN. MSL-1 AND
CC MSL-3 ARE CO-LOCALIZED ON THE X CHROMOSOME. EACH OF THE MSL
CC PROTEINS REQUIRES ALL THE OTHER MSLs FOR WILD-TYPE X-CHROMOSOME
CC BINDING.
CC -1- SUBUNIT: MSL-1 SEEMS TO FORM A TIGHT COMPLEX WITH MSL-2.
CC -1- SUBCELLULAR LOCATION: NUCLEAR; MSL-1 IS ASSOCIATED WITH HUNDREDS
CC OF DISCRETE SITES ALONG THE LENGTH OF THE X CHROMOSOME IN MALES
CC AND NOT IN FEMALES, AND IS ALSO ASSOCIATED WITH 10-20 AUTOSOMAL
CC SITES IN FEMALES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: I42514; AAA8918.1; -
DR FLYBase: FBgn0005617; msl-1.
KW Nuclear protein.
FT CONFLICT 188 193 PLPPAA -> HCHLLP (IN REF. 1).
FT CONFLICT 492 492 L -> S (IN REF. 1).
FT CONFLICT 670 670 I -> M (IN REF. 1).

```

SEQ	SEQUENCE	1039 AA:	117412 MM:	4739E9B5EF6E9P14 CRC64:
Query Match	5.6%;	Score 116.5;	DB 1;	Length 1039;
Best Local Similarity	22.2%;	Pred. No. 7;		
Matches	82;	Conservative	55;	Mismatches 141; Indels 91; Gaps 19;
QY	46	DIIVSDDE-----VLEVADPVEVPVAPRLPAPAKP--EQDSDSDSGAAGPAG-APR	95	
DB	400	DYNATTEELMONGNYLLET-LDPTTEKTCAPSAHSTPHNQKSSYQAEIRKRGNGNRITE	458	
QY	96	TLVRRRRRLDPCGAPVVP--VYSGKVOSSLNLIPDSSSLKLCPSPEDEADLTNSGS	153	
DB	459	KLLQKLPEDMDALAPRLPKWPAKKDKCHELVEPS-----PEVVPKQO-----	503	
QY	154	SPSEEDALPSGSPMKKTKRKCKEKKEMEEPPDDDISPLPPQSSRNKS-----RKHTPAL	209	
DB	504	-PHQEDALVDHNAIKNOL-----EYKPKDLKPKQPKHQEQDQDGLDVRVPEQ	550	
QY	210	OKLREVNRKLDDKSLCSLPKQHSFALQSTDEVVLVESGVPILPOSSRLFTLKIRCADLV	269	
DB	551	EDVRVQKTEKLRQPEDAPK-HLPKA-----VAPKVTKTSRSTPLKANTADI-	598	
QY	270	RLPVMKSEPLQVNVDMHNHLG-----VSPNRLILLFGESLSP-----TATPTSTKL	317	
DB	599	-----KDAPOAKVYI---ANHOSKTQTPDPVKTORLQVAKIROYEMHPDMRGSSAPSDIR-	649	
QY	318	GVADIIDCVLASSSEATSETSOELRLRYOGKKEKHOMLEISLSPDSPLKY-----LMSHY	371	
DB	650	-KQKVVDPV-----STPERKTKIKSKILV--NDKKTTSFSSQSPDQEDIVETVRRRLAEHL	702	
QY	372	EPAMGLSGH	380	
DB	703	KKELLSOSH	711	
RESULT	11			
LDS_DROME				
ID	LDS_DROME	STANDARD:	PRT:	1061 AA.
AC	P34739;			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	01-FEB-1994 (Rel. 28, Last annotation update)			
DE	PROBABLE HELICASE LODESTAR.			
GN	LDS.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephyridiidae; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92009170; PubMed=1916263;			
RA	Girdham C.G., Glover D.M.;			
RA	Bork P., Koonin E.V.;			
RT	"An expanding family of helicases within the 'DEAD/H' superfamily.";			
RT	in <i>lodestar</i> , a <i>Drosophila</i> gene encoding a putative nucleoside			
RT	triphosphate-binding protein.";			
RL	Genes Dev. 5:1786-1799(1991).			
RN	[2]			
RP	CONCEPTUAL TRANSLATION.			
RX	MEDLINE=93181281; PubMed=8382805;			
RA	Bork P., Koonin E.V.;			
RA	"An expanding family of helicases within the 'DEAD/H' superfamily.";			
RL	Nucleic Acids Res. 21:751-752(1993).			
CC	-1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).			
CC	-1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.			
CC	-1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION. A FRAMESHIFT HAD TO			
CC	BE INTRODUCED IN POSITION 946 TO PRODUCE THIS ORF.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X62629; CAA44496.1; ALT_FRAME.
DR PIR: A40580; A40580.
DR FlyBase: FBgn0002542; lds.
DR InterPro: IPR000330; -.
DR InterPro: IPR001650; -.
DR Pfam: PF00176; SNE2_N; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Nuclear protein; Helicase; ATP-binding.
KW NP_BIND 465 472 ATP (POTENTIAL).
FT SITE 603 606 DEAH BOX.
FT SEQUENCE 1061 AA; 118189 MW; E82AA64254342B80 CRC64;

Query Match 5.6%; Score 116.5; DB 1; Length 1061;
Best Local Similarity 22.2%; Pred. No. 7.2; Mismatches 144; Indels 129; Gaps .20;
Matches 92; Conservative 50;

OY 29 PRARSPARLIPDTVLVLDVSDSDSELEVLADPVFVPARLP-APAKPEDSDSDSE--- 84
   | | | | | : | | | | | : | | | | | : | | | | |
Db 94 PSARITKRRKPL-----GIPSDSEDEDEDELRALSPSTRMSITGVRRQDLSDDSELEY 147
   | | | | | : | | | | | : | | | | | : | | | | |
OY 85 -GAAEGPAGAPRTLVRRRRRLRLLDPGEAPVVPVY-----SGKVGSSLN-----LIP 129
   | | | | | : | | | | | : | | | | | : | | | | |
Db 148 SDEVEGEPTLEAP-----TAAEA-VYPGYTTQFAGINQNDLHSTIGAADSEYLD 193
   | | | | | : | | | | | : | | | | | : | | | | |
OY 130 DNSSLILKLCSPSPEDADALITNGSSPSRDDALP-----SGSWRRKKLKK----- 173
   | | | | | : | | | | | : | | | | | : | | | | |
Db 194 DSSGSDVILILSKKEPPIELSS---TDDDATTKENMSGPFEPKSKSLSPSSAGASV 249
   | | | | | : | | | | | : | | | | | : | | | | |
OY 174 -KCEK--EKKKEEPPDDISPLPQPSRRNKSRSK-----HTRALOKLREVNRLDL 222
   | | | | | : | | | | | : | | | | | : | | | | |
Db 250 YKTSNLSQPTIQAVLKKQTSAPAAARRSRIRKSEDDKVVSYQVYDEEMKRLAEKRQVSDA 309
   | | | | | : | | | | | : | | | | | : | | | | |
OY 223 RSCLSPKQHSPPALOSTDDEVVLVGVPLPOSSRLFTIKIRCRADLVRLPYRMSEPLQNV 282
   | | | | | : | | | | | : | | | | | : | | | | |
Db 310 EKLFEPKVAHKLDPKGS-----QIMKRIDLTLRELAMDE--QMI 345
   | | | | | : | | | | | : | | | | | : | | | | |
OY 283 VDHMANHLGVSPNRILLLFGESELSPTLPSTPLKLGVDIIDCVVLASSSEPTSTQGLR 342
   | | | | | : | | | | | : | | | | | : | | | | |
Db 346 SALRYQOSNVPAVRVY---KPTLPPRAPPS-----IDTLDMDELSAVNIEIKIPVY 392
   | | | | | : | | | | | : | | | | | : | | | | |
OY 343 LRVQGS-----KEKHGMLEISLSP---DSP--LKV-LMSHYEAM 375
   | | | | | : | | | | | : | | | | | : | | | | |
Db 393 TGAQGMATFNNOKALITLESKDLHVSLEDDPGEVLAEDPVGLKVMIMNHQKHAL 447
   | | | | | : | | | | | : | | | | | : | | | | |

RESULT 12
TCF8_MESAU STANDARD: PRT: 1043 AA.
ID TCF8_MESAU
AC Q60542;
DC 01-OCR-2000 (Rel. 40, Created)
DT 01-OCR-2000 (Rel. 40, Last sequence update)
DT 01-OCR-2000 (Rel. 40, Last annotation update)
DE TRANSCRIPTION FACTOR 8 (ZINC FINGER PROTEIN B2P).
GN TCF8 OR B2P.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Insulinoma;
RX MEDLINE=95021206; PubMed=7935395;
RA Franklin A.J., Jettion T.L., Shelton K.D., Magnuson M.A.;
RT "B2P, a novel serum-responsive zinc finger protein that inhibits gene
transcription."
MOl. Cell. Biol. 14:6773-6788(1994).
CC -i- FUNCTION: ACTS AS A TRANSCRIPTIONAL REPRESSOR. BINDS TO E-BOX

```



```

      823 LAPAKESPRKCAAPTGGCKTGPSSAAQAKODDSCGSSSESDSDGEAPAAVTSAOVIRPPL 882
      153 --SSPSDDALPSSGSPWR-----KKLRKCKEKEKKMEFPDODISPLPQ---PSSRN 200
      883 IEPDNPSPAGPAPYPPAQAQAASGTPPKARASESTARSSSSSEDEEDYIPATQCLTPGIRT 942
      201 K-----SRKHTALQKLRVKNKRLQDRLSCLSPKQHSQSPALQSTDEYVVEGP 249
      943 NVVTMPRAHPRIAKKAMAGASSKESRISD-----GKKQEGPATQ-----VSKKNP 990
      250 V-LPQSSRLFTLKTRC-RADIVRLPVRMSEPLQNVVDHMANHLGVSPPRILLFGESELS 307
      991 ASLFLTQA--ALKVLAQKASEAPQPVARTQPSGCVDSAVGTLPATSPQ-----STSVQ 1041
      308 PTAPSTLKLGVADIIIDCVVLASSEATEPTSQE 340
      1042 AKGTNKLKPKLPEVQATKAPRESSDSESDS 1074

```

RESULT 15

```

DYN_A_RAT
ID DYN_A_RAT STANDARD; PRT; 1280 AA.
AC P28023;
DT 01-AUG-1992 (Rel. 23, Created)
DT 15-JUL-1998 (Rel. 35, Last sequence update)
DE DYNACTIN, 150 KDA ISOFORM (150 KDA DYNEIN-ASSOCIATED POLYPEPTIDE)
DE (DP-150) (DAP-150) (P150-GLUED).
GN DCTN1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX MEDLINE=91260877; PubMed=1828535;
RA Holzbaur E.L.F., Hammarback J.A., Paschal B.M., Kravitt N.G.,
RA Pfister K.K., Viallee R.B.;
RT "Homology of a 150k cytoplasmic dynein-associated polypeptide with
RT the Drosophila gene Glued.";
RL Nature 351:579-580(1991).
RN [2]
RP REVISIONS.
RA Holzbaur E.L.F., Hammarback J.A., Paschal B.M., Kravitt N.G.,
RA Pfister K.K., Viallee R.B.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: REQUIRED FOR THE CYTOPLASMIC DYNEIN-DRIVEN RETROGRADE
CC MOVEMENT OF VESICLES AND ORGANELLES ALONG MICROTUBULES. DYNEIN-
CC DYNACTIN INTERACTION IS A KEY COMPONENT OF THE MECHANISM OF AXONAL
CC TRANSPORT OF VESICLES AND ORGANELLES.
CC -!- SUBUNIT: LARGE MACROMOLECULAR COMPLEX OF AT LEAST 10 COMPONENTS.
CC P150(GLUED) BINDS DIRECTLY TO MICROTUBULES AND TO CYTOPLASMIC
CC DYNEIN.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: CONTAINS 1 CAP-GLY DOMAIN.
CC -!- SIMILARITY: STRONG. TO OTHER SPECIES DYNACTIN 150 KDA SUBUNIT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X62160; CAA44091.1; -
CC PIR; S16129; S16129.
CC InterPro: IPR000938; -
CC Pfam; PF01302; CAP_GLY; 1.
CC PROSITE; PS00845; CAP_GLY_1; 1.
CC Motor protein; Microtubules; Dynein; Coiled coil; Cytoskeleton.
CC

```

```

FT DOMAIN 48 90 CAP-GLY.
FT DOMAIN 157 184 SER-RICH.
FT DOMAIN 214 513 COILED COIL (POTENTIAL).
FT DOMAIN 942 1048 COILED COIL (POTENTIAL).
FT DOMAIN 1184 1213 COILED COIL (POTENTIAL).
SO SEQUENCE 1280 AA; 141929 MW; C9348CF129F4FF5C CRC64;

```

Query Match 5.4%; Score 113.5; DB 1; Length 1280;
 Best Local Similarity 22.6%; Pred. No. 14; Matches 87; Conservative 61; Mismatches 144; Indels 93; Gaps 17;

```

      71 APAPEDSDSDSGAEGPAGAPRTLVRRRRRLDPGEAPVY-----PYYS 118
      105 SPETP--DSSASKILKREGADAAKT-----SKLGLPKPKAPTKATTTRPKPTRPAST 158
      119 GKVOSSLNLPDNS-SLKLCSPSEDEAD-----LNSGSSPEDALPSGS 165
      159 GVAGPSSSLGPGSGASAGELSSSESPSPAPQPLAP11PPPALTPGAP--PLSPSS 214
      166 PMRKKLRKCKEKEKKMEFPDQDISPLPQPSRNKRS-----KHTALQKLRVKNRLQ 220
      215 KEEGLRDQVADLEKLE-----TLRLKRSDEKAKLKEKHKIQLEVGQEMKSKMQ 266
      221 ----DLRSCLSPKQHSQPALQSTD-----DEVVVEGVPVLPQSSRLFTLKIRCADLV 269
      267 EQADLQRLRKEAKEAKEAKEAKERYMEADTADAEIMATLDKE-----MAEERAEVL 320
      270 RLPVPMSEPLQNVVDHMANHLGVSPNRILLFGSELSPTATPS----TLKLGVADIIIDC 325
      321 QOEV---EALKERYDELTTDEILIKAEI-----BEKSGDGAASSYOLKQLEONARLKDA 372
      326 VVLASSSEATEPTSGELRLRVQGKEKHOKLEISLSPDSPLYKLYMSHYEAMLSGHLKLSFF 385
      373 LVRRDLSSSKQEHVLOKLEKKNQDELEVYRQORERLOELQAESTI----- 422
      386 FDGTLKSGKELPADLGESGDLIEV 410
      423 -DELK---EQVDALGAE--EMVEM 441

```

Search completed: September 11, 2001, 08:56:27
 Job time: 416 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 11, 2001, 08:48:26 ; Search time 85.95 Seconds
(without alignments)
1063.422 Million cell updates/sec

Title: US-09-192-611-2

Perfect score: 2099
Sequence: 1 MAEPLRGKGRPSRGGRARR.....GKELPADLGESGDLIEWG 412

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2803329 seqs, 221847457 residues

Total number of hits satisfying chosen parameters: 2803329

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_AA_Main:*

1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep:*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep:*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep:*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep:*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep:*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep:*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep:*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep:*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep:*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep:*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep:*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep:*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep:*
21: /cgn2_6/ptodata/2/paa/US097_COMB.pep:*
22: /cgn2_6/ptodata/2/paa/US098_COMB.pep:*
23: /cgn2_6/ptodata/2/paa/US099_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2099	100.0	412	11	US-08-755-592-6
2	2099	100.0	412	11	US-08-755-592A-6
3	2099	100.0	412	15	US-09-175-254-4
4	1399.5	66.7	419	15	US-09-175-254-3
5	154.5	7.4	1279	1	PCT-US01-04098A-1631
6	154.5	7.4	1301	1	PCT-US01-04098A-3599
7	145.5	6.9	622	20	US-09-641-377-505
8	145.5	6.9	628	20	US-09-641-377-506
9	141	6.7	530	20	US-09-641-377-507
10	133.5	6.4	725	16	US-09-252-991A-23752

11	132	6.3	1708	21	US-09-733-089-18072	Sequence 18072, A
12	132	6.3	1708	22	US-09-816-660-18072	Sequence 18072, A
13	130.5	6.2	1488	21	US-09-733-089-18070	Sequence 18070, A
14	130.5	6.2	1488	20	US-09-816-660-18070	Sequence 18070, A
15	128.5	6.1	755	22	US-09-619-049-240	Sequence 240, App
16	128.5	6.1	755	23	US-60-161-932-1882	Sequence 1882, Ap
17	128.5	6.1	755	23	US-60-167-217-4680	Sequence 4680, Ap
18	128.5	6.1	755	23	US-60-171-622-372	Sequence 372, App
19	128.5	6.1	755	23	US-60-173-464-3782	Sequence 3782, Ap
20	128.5	6.1	755	23	US-60-173-464-3782	Sequence 4587, Ap
21	128.5	6.1	755	23	US-60-191-637-4587	Sequence 3621, Ap
22	128.5	6.1	2024	23	US-60-167-217-5735	Sequence 5735, Ap
23	128.5	6.1	2024	23	US-60-173-464-4625	Sequence 4625, Ap
24	127.5	6.1	1167	23	US-60-167-217-56	Sequence 51, Appl
25	127.5	6.1	1167	23	US-60-173-464-51	Sequence 63, Appl
26	127.5	6.1	1167	23	US-60-191-637-63	Sequence 51, Appl
27	127.5	6.1	1167	23	US-60-191-681-51	Sequence 2935, Ap
28	127.5	6.1	1408	23	US-60-167-217-2935	Sequence 2410, Ap
29	127.5	6.1	1408	23	US-60-173-464-2417	Sequence 2910, Ap
30	127.5	6.1	1408	23	US-60-191-637-2910	Sequence 40178, A
31	127.5	6.1	442	18	US-09-417-507-40178	Sequence 396, App
32	127	6.1	1368	20	US-09-619-049-396	Sequence 7686, Ap
33	127	6.1	1368	23	US-60-167-217-7686	Sequence 633, App
34	127	6.1	1368	23	US-60-171-627-633	Sequence 6151, Ap
35	127	6.1	1368	23	US-60-173-464-6151	Sequence 7599, Ap
36	127	6.1	1368	23	US-60-191-637-7599	Sequence 5928, Ap
37	127	6.1	1368	23	US-60-191-681-5928	Sequence 28, Appl
38	127	6.0	676	1	PCT-US01-04703-28	Sequence 29, Appl
39	126.5	6.0	717	1	PCT-US01-04703-29	Sequence 602, App
40	126.5	6.0	664	23	US-60-212-664-602	Sequence 862, App
41	125.5	5.9	518	1	PCT-US00-05883-862	Sequence 5837, App
42	124.5	5.9	555	1	PCT-US00-26524B-5837	Sequence 307, App
43	124.5	5.9	1779	23	US-60-245-228-307	Sequence 3865, Ap
44	124.5	5.9	1191	1	PCT-US01-04098A-3865	
45	124	5.9				

ALIGNMENTS

RESULT 1

US-08-755-592-6

Sequence 6, Application US/08755592

GENERAL INFORMATION:

APPLICANT: Glimcher, Laurie H.

APPLICANT: Ho, I-Cheng.

TITLE OF INVENTION: Methods for Regulating T Cell Subsets by

NUMBER OF INVENTIONS: 6

MODulating Transcription Factor Activity

CORRESPONDENCE ADDRESS:

ADDRESSER: LAHIVE & COCKFIELD

STREET: 60 State Street, suite 510

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/755,592

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/636,602

FILING DATE: 23-APR-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Decont, Giulio A., Jr.

REGISTRATION NUMBER: 31,503

REFERENCE/DOCKET NUMBER: HUT-021CP

TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 412 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-755-592A-6

Query Match 100.0%; Score 2099; DB 11; Length 412;
Best Local Similarity 100.0%; Pred. No. 1.4e-153;
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEPLRGGRGPRSGRGARRARAGRCPRAROSPAPRLIPDTVLVDVSDSEVLEVAD 60
DB 1 MAEPLRGGRGPRSGRGARRARAGRCPRAROSPAPRLIPDTVLVDVSDSEVLEVAD 60
QY 61 PVEVPVAPRLPAPAKPEODSDSDSEGAAGPAGAPRTLVRRRRRLDPPGEAPVVPVYSGK 120
DB 61 PVEVPVAPRLPAPAKPEODSDSDSEGAAGPAGAPRTLVRRRRRLDPPGEAPVVPVYSGK 120
QY 121 VOSSLNLIIPNSSLILKICPSEPEDEADLTNSGSSPSEDALPGSPWRKKLRKKCEKEEK 180
DB 121 VOSSLNLIIPNSSLILKICPSEPEDEADLTNSGSSPSEDALPGSPWRKKLRKKCEKEEK 180
QY 181 KMEFFPDDISPLPQPSRNRKSRKHTALOKLREVNKRLODLRSCLSPKQHSALOSTD 240
DB 181 KMEFFPDDISPLPQPSRNRKSRKHTALOKLREVNKRLODLRSCLSPKQHSALOSTD 240
QY 241 DEVVLVGPVLPQSSRLFTLKIRCADLVRLPYRMSEPLQNVVDHMANHILGVSPNRILL 300
DB 241 DEVVLVGPVLPQSSRLFTLKIRCADLVRLPYRMSEPLQNVVDHMANHILGVSPNRILL 300
QY 301 FGESELSPTATPTSTLKIGVADIIDCVVLASSSEATETSQELRLRVQGEKHKQMLEISLSP 360
DB 301 FGESELSPTATPTSTLKIGVADIIDCVVLASSSEATETSQELRLRVQGEKHKQMLEISLSP 360
QY 361 DSPKLVMISHYEAMGLSGHKLSFFPDGTRKLSGKELPADLGLESGLIEYWG 412
DB 361 DSPKLVMISHYEAMGLSGHKLSFFPDGTRKLSGKELPADLGLESGLIEYWG 412

RESULT 2

US-08-755-592A-6
Sequence 6, Application US/08755592A
GENERAL INFORMATION:
APPLICANT: Glimcher, Laurie H.
APPLICANT: Ho, I-Cheng.
TITLE OF INVENTION: Methods for Regulating T Cell Subsets by
TITLE OF INVENTION: Modulating Transcription Factor Activity
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street, suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/755,592A
FILING DATE: 25-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/636,602
FILING DATE: 23-APR-1996

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kara, Catherine J.
REGISTRATION NUMBER: 41,106
REFERENCE/DOCKET NUMBER: HUI-021CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 412 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-755-592A-6

Query Match 100.0%; Score 2099; DB 11; Length 412;
Best Local Similarity 100.0%; Pred. No. 1.4e-153;
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEPLRGGRGPRSGRGARRARAGRCPRAROSPAPRLIPDTVLVDVSDSEVLEVAD 60
DB 1 MAEPLRGGRGPRSGRGARRARAGRCPRAROSPAPRLIPDTVLVDVSDSEVLEVAD 60
QY 61 PVEVPVAPRLPAPAKPEODSDSDSEGAAGPAGAPRTLVRRRRRLDPPGEAPVVPVYSGK 120
DB 61 PVEVPVAPRLPAPAKPEODSDSDSEGAAGPAGAPRTLVRRRRRLDPPGEAPVVPVYSGK 120
QY 121 VOSSLNLIIPNSSLILKICPSEPEDEADLTNSGSSPSEDALPGSPWRKKLRKKCEKEEK 180
DB 121 VOSSLNLIIPNSSLILKICPSEPEDEADLTNSGSSPSEDALPGSPWRKKLRKKCEKEEK 180
QY 181 KMEFFPDDISPLPQPSRNRKSRKHTALOKLREVNKRLODLRSCLSPKQHSALOSTD 240
DB 181 KMEFFPDDISPLPQPSRNRKSRKHTALOKLREVNKRLODLRSCLSPKQHSALOSTD 240
QY 241 DEVVLVGPVLPQSSRLFTLKIRCADLVRLPYRMSEPLQNVVDHMANHILGVSPNRILL 300
DB 241 DEVVLVGPVLPQSSRLFTLKIRCADLVRLPYRMSEPLQNVVDHMANHILGVSPNRILL 300
QY 301 FGESELSPTATPTSTLKIGVADIIDCVVLASSSEATETSQELRLRVQGEKHKQMLEISLSP 360
DB 301 FGESELSPTATPTSTLKIGVADIIDCVVLASSSEATETSQELRLRVQGEKHKQMLEISLSP 360
QY 361 DSPKLVMISHYEAMGLSGHKLSFFPDGTRKLSGKELPADLGLESGLIEYWG 412
DB 361 DSPKLVMISHYEAMGLSGHKLSFFPDGTRKLSGKELPADLGLESGLIEYWG 412

RESULT 3

US-09-175-254-4
Sequence 4, Application US/09175254
GENERAL INFORMATION:
APPLICANT: Hong Zhou
APPLICANT: Jiaogiao Zhao
APPLICANT: Derong Liu
TITLE OF INVENTION: NIP45 HUMAN HOMOLOG
FILE REFERENCE: PHM.70281
CURRENT APPLICATION NUMBER: US/09/175,254
CURRENT FILING DATE: 1998-10-20
EARLIER APPLICATION NUMBER: GB 9722388.7
EARLIER FILING DATE: 1997-10-24
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 412
TYPE: PRT
ORGANISM: Mus musculus
US-09-175-254-4

Query Match 100.0%; Score 2099; DB 15; Length 412;


```
OY 202 SRKHTALOK-----LREYNKRLOD-----LRSLSPKOHSPALOSTDDEVVL 245
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 733 SEQVOAPARGTIGRYVPGRSSEKHSPPDACSVDYSSSLSPER--PDESESTEPIS 790
OY 246 VEGPVLPOSSRLFTLIRCRADVLRLPVNMSBPLQNVNDMANHLGVSPNRILLFGESE 305
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 791 VDG-----ISSDLEPAEGDEEEBEGMGCP-----YGLQE 822
OY 306 LSPFATP-----STIKLGADIDICVYVLAASSSEAEETQDELRLRQGEKQKMLE 355
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 823 GSP-QPDPDQFLKQHETLTLASGAAPCAP-VQYPERSESHSISRFLLOVQTRPLREP-- 878
OY 356 ISLSPDPLKVLKSHYEAMGLSGHK 381
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 879 ----SPSSSLALMSRPAQVPAASGEQ 901

RESULT 6
PCT-US01-04098A-3599
; Sequence 3599, Application PC/TUS0104098A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-029
; CURRENT APPLICATION NUMBER: PCT/US01/04098A
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: Not Yet Assigned
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 09/728,422
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 09/693,325
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/663,561
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 09/654,936
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/620,325
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/598,075
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 3960
; SOFTWARE: Custom
; SEQ ID NO 3599
; LENGTH: 1301
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-04098A-3599

Query Match 7.4%; Score 154.5; DB 1; Length 1301;
Best Local Similarity 23.5%; Pred. No. 0.025;
Matches 105; Conservative 48; Mismatches 164; Indels 129; Gaps 21;

OY 6 RGGGPRSGRGARRARGRCPRARQSPARLIPDTVLVDLYSDSEVELEVADEPVEVP 65
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 537 RORGGKQGGSPSSPQASG-----PNRHQAPSMLSGPCA--LSSDSDKEGEDSGTEEBELP 568
OY 66 -----VARLPAPAKPEODSDSDSEGAEE-----GPAGAPRTLVRRRRRL 106
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 589 ALPVLAKSTKKALASVSPALPRLPSLSHWEMSRQDESYGFLDPAPANPGP---RRRGKRV 645
OY 107 DPEGAIPVYSGKVOSSLNLIIPDNSSLKLCPS--EP-EDEADLTNNGSSSPSEDALPS 163
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 646 QPG-----VELSVSMIDLRLQLETLAPSLQDPQDSLAIIPSGPRKHQGEALLET 694
OY 164 G-----SPWRKTLRKKC-----EKKE-----KMEEFPPDDISPLPQPS---SRNK 201
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 695 SLTSGNEKPPRQASQPCSYPIHRIILLSDQEVGAODLEPAPIEDGIVPEPDSNPMTMDT 754
```

```
OY 202 SRKHTALOK-----LREYNKRLOD-----LRSLSPKOHSPALOSTDDEVVL 245
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 755 SEQVOAPARGTIGRYVPGRSSEKHSPPDACSVDYSSSLSPER--PDESESTEPIS 812
OY 246 VEGPVLPOSSRLFTLIRCRADVLRLPVNMSBPLQNVNDMANHLGVSPNRILLFGESE 305
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 813 VDG-----ISSDLEPAEGDEEEBEGMGCP-----YGLQE 844
OY 306 LSPFATP-----STIKLGADIDICVYVLAASSSEAEETQDELRLRQGEKQKMLE 355
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 845 GSP-QPDPDQFLKQHETLTLASGAAPCAP-VQYPERSESHSISRFLLOVQTRPLREP-- 900
OY 356 ISLSPDPLKVLKSHYEAMGLSGHK 381
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 901 ----SPSSSLALMSRPAQVPAASGEQ 923

RESULT 7
US-09-641-377-505
; Sequence 505, Application US/09641377
; GENERAL INFORMATION:
; APPLICANT: WIEMANN, STEFAN
; APPLICANT: GASSENHUBER, JOHANN
; APPLICANT: TAMPE, JENS
; TITLE OF INVENTION: HUMAN DNA SEQUENCES
; FILE REFERENCE: 087100/0106
; CURRENT APPLICATION NUMBER: US/09/641,377
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: 60/149,499
; PRIOR FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: 60/156,503
; PRIOR FILING DATE: 1999-09-28
; NUMBER OF SEQ ID NOS: 1793
; SOFTWARE: Patentlin Ver. 2.1
; SEQ ID NO 505
; LENGTH: 622
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Database comparison
; OTHER INFORMATION: sequence used for homology purposes
US-09-641-377-505

Query Match 6.9%; Score 145.5; DB 20; Length 622;
Best Local Similarity 22.1%; Pred. No. 0.042;
Matches 95; Conservative 63; Mismatches 166; Indels 105; Gaps 20;

OY 4 PLRGGRPSRSGRGARRARGRCPRARQSPARLIPDTVLVDLYSDSEVE----- 55
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 178 PVEAKSPAEEKSPASVAKSPG-EAKSPAEEKSPADEVSPATVKSPEVEKSPAEEKSVYTK 236
OY 56 --LEVADPVEVPVYARLPAPAKPEODSDSDSEGAEBGAPGAPRTLVRRRRRLIDPEGAPV 113
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 237 SPAEAKSPVEV-----KSPASVAKSPSEAKSPAGAKSPAEEKSPVYAKSPAEEKSPA- 290
OY 114 VPYVSGKVOSSLNLIIPDNSSLKLCPSPEPEDADLTN-----SGSSPSEDA--LP 162
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 291 -PPEAKSPAEEKSPAEEKS-----PAEAKSPAEEKSPVEVKSPEEAKSPVKGAGKSLAE 344
OY 163 SGSPWRKTLRKCKEKEKMEFPDDISPLPQPS--SRNKS-----KHTEALOKLREVN 217
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 345 AKSPEAKSPVKEIKRPPAIVKSPKAKSPKMEAKSPEAKTLVDVKSPEAKTPAKMEAK 404
OY 218 RLDDNLSCSLPKOHOSPALOSTDDEVVLVEGVLPOSSRLFTLKICRADVLPLP--VR 274
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 405 RPADIR---SPEQVKSPEAKEEAKS-----PEKEETRTKEVAPKKEEVKSPVEEVK 451
OY 275 MSBPLQNVNDMANHLGVSPNRILLFGESELSPTAPPSPLKGVADIIDCVYVLAASSSEA 344
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 452 AKEP-----PKKV-----EKEKP-ATPKTE-----VSKKKDEA 480
```



```

; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23752
; LENGTH: 725
; TYPE: prt
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23752

```

Query Match	6.48;	Score 133.5;	DB 16;	Length 725;
Best Local Similarity	28.08;	Pred. No. 0.44;		
Matches	90;	Conservative 32;	Mismatches 128;	Indels 71;
				Gaps 20;

OY	8	RGPR--SNGSGARRARARACRCRACROSPARLLIPDVLVLDVSDSDSEVLEVAADPEVPE	65
Db	287	KRPRSPGPGKILSRSLRR--FRAROEQACPRLLAORLPBGRGQPPQRYAAQRPLP	344
OY	66	VARLPAPA--KPEODSDSEGAAGPACAPRTLVRRRRRLLDPGCAV-----VPVY	117
Db	345	AAGRPSPGGLERPLAAHNGDSRGAA--AAGAG---NHRLRLCEPRLPAGLPQPLPGH	399
OY	118	S-GKYOSLNLIPNSSLLKLCSEPEPEADLTWSSGSP-----SEPDALPQSG--PWK	169
Db	400	ALGGLVSS-----PADAA-EAHHPRHAERGGDRTGRCGNPLRLRSGGALPLQSGRCGR	454
OY	170	K---LRKKCEK---EKKMEEP---DODISPLPQPSR-----NKSRTKEA	208
Db	455	QOOLLVTRCLTGLGKELTCHNRHPRGRRGAGADLPERRRSRPAAGHRLPVSOGTGAALTAG	514
OY	209	LQKLRVVKRYLQDLRSCSLSPKQOHPALQSDTDEVLVCEP---VLPOSSRLFLTKR	265
Db	515	AQDPQRLODPQPPRRRWLQRAAHNGP-----QCPGADIRCPARL-CRKHKRAR	561
OY	266	ADLVRLPV---RMSEPLQNV	282
Db	562	PEDEPRSETPRRHAGEOLOLV	582

```

RESULT 11
US-09-733-089-18072
; Sequence 18072, Application US/09733089
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kowalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; APPLICANT: Wu, Wei
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; TITLE OF INVENTION: Transcription In Plants
; FILE REFERENCE: 38-21(13300)D
; CURRENT APPLICATION NUMBER: US/09/733,089
; CURRENT FILING DATE: 2000-12-11
; PRIOR APPLICATION NUMBER: US 09/474,435
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: US 09/654,617
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 09/620,392
; PRIOR FILING DATE: 2000-07-19
; NUMBER OF SEQ ID NOS: 24143
; SEQ ID NO 18072
; LENGTH: 1708
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-733-089-18072

```

Query Match	6.3%	Score 132;	DB 21;	Length 1708;
Best Local Similarity	25.1%;	Pred. No. 2.1;		
Matches	67;	Conservative	27;	Mismatches 113; Indels 60; Gaps 9;
QY	10	PRSGGRCARRA--RGAGRCPRANO--SPARKLPDTVLVDVSDSEVLEAVADPEVPV	66	
	*	: : : : : :		

```

Db 635 PPSRDAEERREAMEGCAAREBNVYDVEPILGDDDAKVDPQTSRPMBEITSAEVCVEV 694
Qy 67 ARLPAPAKPEODSDSDEGAEGPAGABRTLVRRRRRLDPEGAPVYVYSGKVOSSLN 126
Db 695 SEASPLGAEED-----EGKPGSTRNLRKLRKTTTIDSAAEHGAESDEASCOE 745
Qy 127 LIPDNSSLKLCP-----SEPEDEADL-----TNGSGSPSEDDALPSCSPMWRKL 171
Db 746 LAPCRSRRLRCKPPIRKSRSKTPGCKHDLGKCSEDDVFNANSSSSQDDBNCSAPRGRSKL 805
Qy 172 RKCKEERKEKMEEPDODISLPQPSRHNKSRRKHTALQKLRVNRKRLDOLR----- 223
Db 806 RMR-----LCPRNISYADDSDTDIDNRK---KIVPCRRTSKRLQEKRAKADVSEE 852
Qy 224 -----SCLSPKQHSPALQSTDEDEV 244
Db 853 SYTEASSCMLSTS-----SSSDELL 873

```

```

? RESULT 12
? US-09-816-660-18072
? Sequence 18072, Application US/09816660
? GENERAL INFORMATION:
? APPLICANT: Dotson, Stanton B.
? APPLICANT: Kovall, David K.
? APPLICANT: Liu, Jingdong
? APPLICANT: Lutfiyya, Linda L.
? APPLICANT: McIninch, James
? APPLICANT: Wu, Wei
? TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
? TITLE OF INVENTION: Transcription In Plants
? FILE REFERENCE: 38-21(15300)D
? CURRENT APPLICATION NUMBER: US/09/816,660
? CURRENT FILING DATE: 2001-03-26
? PRIOR APPLICATION NUMBER: US 09/474,435
? PRIOR FILING DATE: 1999-12-28
? PRIOR APPLICATION NUMBER: US 09/654,617
? PRIOR FILING DATE: 2000-09-05
? PRIOR APPLICATION NUMBER: US 09/733,089
? PRIOR FILING DATE: 2000-12-11
? PRIOR APPLICATION NUMBER: ) US 09/684,016
? PRIOR FILING DATE: 2000-10-10
? PRIOR APPLICATION NUMBER: US 09/620,392
? PRIOR FILING DATE: 2000-07-19
? NUMBER OF SEQ ID NOS: 24143
? SEQ ID NO 18072
? LENGTH: 1708
? TYPE: PRT
? ORGANISM: Oryza sativa
? US-09-816-660-18072

```

Query Match	Similarity	6.3%	Score 132	DB 22	Length 1708
Best Local Match	67	Conservative	25.1%	Pred. No. 2.1	Mismatches 113; Indels 60; Gaps 9
QY	10	PSRGGRGARRA--RGAGRCPRARQ--SEARLLPDTVLVDLSDESDLEVLEVADPEVEPV	66		
Db	635	PSRPAAGRRRAMEGARARGENVVVEPIKLCDDDAKVDPOTKSPNEITTSAEVCEVT	694		
QY	67	ALPLPAKREOOSDSDSEBGAEGPAGPTVLVRRRRRLLDGCEAPVPVPEYSSKVVOSLNT	126		
Db	695	SEASPLGABEE-----BGRPESTRNLRLKRTITITDSAAAEHGAESDDEAFSCOE	745		
QY	127	LIPDNSSLKLKLP-----SEPDEADL-----TNSGSSPSEDDALPSCSPWRKL	171		
Db	746	LAPCRSRKRLRCKPTKRFRSKPTGKRLDGLGSEDDVFAVNSSSSODDENCASAPRRGSRKL	805		
QY	172	RKKCKEKKMKMEFPQDLSPLPQPSRKNSKHTFALQKLEVNRRLDGLR-----	223		
Db	806	RNR-----LCGRNLTSYADDDSDTIDNRK-----KIVPCRTSRKRLDCKRAADVSEE	852		
QY	224	-----SCLSPKQHQSPALQSTDDDEV	244		


```

: LOCATION (31)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
:
: NAME/KEY: SITE
: LOCATION (126)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
:
: NAME/KEY: SITE
: LOCATION (129)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
:
: NAME/KEY: SITE
: LOCATION (135)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
:
: NAME/KEY: SITE
: LOCATION (136)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
:
: US-09-758-472-5605

```

Query Match	9.68	Score	201.5	DB 5	Length	146			
Best Local Similarity	41.78	Pred. No.	5.6e-06						
Matches	58	Conservative	9	Mismatches	57	Indels	15	Gaps	3

```

0y      4 PLRGSPRSRGGRGARAR-----GARGCPRAGQSPARLP---DYLVDLYSDSDEE 54
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
db      1 PTRPGCGGEXGPLVRRXRCRPRGSGRLGFXPASSGPAVSIPIGHAGMSVDLVTDSDEE 60

```

QY 55 VLEV-----ADPVEYFVARLEPAKPEQSDSDSEGADEGPAGAPRTLVRRRRRLLDP 108
:||| ||||| :||| :|||
Db 61 ILEVATARGADEVEYEPPEPPGVASRDNSNSEGEDRRPADPRGTGSADGGRLVLPD 120

```
QY 109 GEAPVVPVYSGKVQSSNL 127
    ||||: || |||: |
Db 121 GEAPLPVXSGKLKXXFAL 139
```

```

1  RESULT      6
2  US-09-758-472-9144
3  : Sequence 9144, Application US/09758472
4  : GENERAL INFORMATION:
5  : APPLICANT: Rosen et al.
6  : TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
7  : FILE REFERENCE: PH001
8  : CURRENT APPLICATION NUMBER: US/09/758, 472
9  : PRIOR FILING DATE: 2001-01-11
10 : PRIOR APPLICATION NUMBER: 60/179,065
11 : PRIOR FILING DATE: 2000-01-31
12 : PRIOR APPLICATION NUMBER: 60/180,628
13 : PRIOR FILING DATE: 2000-02-04
14 : NUMBER OF SEQ. ID NOS: 9632
15 : SOFTWARE: patentIn Ver. 2.0
16 : SEQ. ID NO 9144
17 :
18 : LENGTH: 78
19 :
20 : TYPE: PRT
21 : ORGANISM: Homo sapiens
22 : US-09-758-472-9144

```

Query Match	7.8%	Score 164	DB 5	Length 78
Best Local Similarity	50.6%	Pred. NO.	0.00048	
Matches 39	Conservative 9	Mismatches 27	IndeIs 2	Gaps 2

```

QY 109 GEARVVPVYSGKVQSSLNLPNSSLKLCPSFEDEADITNSGSSPSEDALPSCSPWR 168
    | : | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 GEPRLFPFTTGRFKSSRLIPDDLSLKLYPPGDDEEALADSSGLYHEGSPSP -GSPWK 61

```

```
QY      169 KKLKKCEKEKKMEEF 185
          ||| | : |||| | |
Db      62  TKLR TK-DKEEKKKTTF 77
```

```

RESULT 7
PCT-US01-14827-8086
; Sequence 8086, Application PC/TUS0114827
; GENERAL INFORMATION:

```

```

APPLICANT: Hyseq, Inc
TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
FILE REFERENCE: 21272-104
CURRENT APPLICATION NUMBER: PCT/US01/14827
CURRENT FILING DATE: 2001-05-16
PRIOR APPLICATION NUMBER: 09/577,408
PRIOR FILING DATE: 2000-05-18
NUMBER OF SEQ ID NOS: 16102
SOFTWARE: Custom
SEQ ID NO 8086
LENGTH: 716
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: DOMAIN
LOCATION: (654)..(706)
OTHER INFORMATION: Ublgqtltn domain proteins domain identified by EMAPRFX.
OTHER INFORMATION: accession number BL00299, p-value=4.420e-18, raw score of 28.8
NAME/KEY: DOMAIN
LOCATION: (401)..(608)
OTHER INFORMATION: Reverse transcriptase domain identified by Pfam, accession
OTHER INFORMATION: name rvt, E-value=2.2e-11, Pfam score of 51.2
PCT-US01-14827-8086

```

Query Match	7.0%;	Score 146;	DB 1;	Length 716;
Best Local Similarity	18.7%;	Pred. No. 0.15;		
Matches	89;	Conservative	73;	Mismatches 173; Indels 142; Gaps 17

```

QY      48  VSDSDSEVLEVADPVEYVVARLEAPAKPEQ--DSDSDSECAAEIAGAPRTLVRRRRRR 104
      ::|  |  :  :  :  |||  |  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      261  INESDSWFFERINKIDRPLARLKKKKKKRKNPIDAIKNDGDIITTPTEIGTTIREYVKH 320

```

```

QY      105 LDDPEAPVYVYSGKQVS--SLNLIIPNSSLKLCKPSEPD-EADITNS-----GSSP 15
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      321 L-----YANKLENLENDKFLDQYTLPRLNQEEVESLNRPITGSEIVALINSLP 363

```

```

Oy 156 SEDDALPSC-----SPWRKRIKKCEEEKKKEEFPDDDISLPQPSSRKKS 20
    :: | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 370 TKKSPGPDGFTTESYORRKEELVPLLKKSQSIKEGILPNSFYELSIILIPK-GRDTT 420

```

```

QY      203 RHH-----FALQKL-----REVN----- 210
      :|      : : :|
Db      429 KKENRPISLNIDAKILINKILANOIOQHRIKLHHDDQVGRIPGMOGWENICKSINVRQH 480

```

```

QY 217 -----KRLQDRSC-----LSPKQHSPA-----LQS 238
      |  ::  |  |  |  |  |
Db 489 RNRATDKNHMTISIDAEKAFDELQOCFEMKTLNKLKPSNHTALAGYGVHOEDSIALLAYD 548

```

QY 239 TDDEVVVEGFPVL PQSSRLFTLIRCADLVRLPYRMSEPLQNVDHMANHLGVSPNRI - 29
| : | | : | : : | : : | : | :
Db 549 LDENHNESOLAAPDSHIIT---SCRKNLSAPESAGSAVHAAP-MHSCASSPPIRYR 60

```

QY      298  ---LLEGESELSPTANPSTLKGADIIDCVVLASSSEATETSQELRLRVGCKEKHOML 35
      | : : | : | : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      605  TRHLCCEAAAEETPVPMANEK-----PTEEVKTENNHNILNKVAGOD-GSYV 65

```

QY 355 ELSLSPDPLVLMSHYEAMGISHKLSFFEDGTKLSEKELPADGLESDLIEMW 4111:
: : || || | | | | | | | | : | | :
Db 652 QFKIKRPTLSKIMKAYCEPRGISMKOIRFEEGGPISTDDKPAOLEMEDEPRTIDVF 708

```

RESULT      8
PCT-US01-14827-14803
: Sequence 14803, Application PC/TUS0114827
: GENERAL INFORMATION:
: APPLICANT: HYSENG, INCV
: TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
: FILE REFERENCE: 21272-104
: CURRENT APPLICATION NUMBER: PCT/US01/14827
: CURRENT FILING DATE: 2001-05-16
: PRIOR APPLICATION NUMBER: 09/577,408
: PRIOR FILING DATE: 2000-05-18

```


QY 29 PRARQSP-----ARLPDTVLVDLVSDSEVLEVADPVEV--PVARLP--APAKPE-Q 77
Db 34 PKTESPEKEPEKSEVKPTDPTLDPKVEEGKEDSAPAPVEVEGVESEKPEEKVAKPESQ 93
QY 78 DSDSDSGAAGGAPGARITLVRRRRRLDLDBGAPVYVYSGKVOSS---LNLIPDNSSL 134
Db 94 PSDRPAEESKVEQAGEP-VAPREDEKAPVEPEKQPEAPEEKAEETPPKQESTPTDKAE 152
QY 135 LKICP-----SEPEDEADLTNSGS--SPSDDALPSGSPRRK 170
Db 153 EYVPEKPEYVNOSEQKVEFPANVEKOTEPPEEKVQAGEPVAPREDEQAPT-APVEPE 211
QY 171 LRKCKEKKKEEPPDDISPLPQPSRRKSRKHTALOKLREVNKRLODLNSCLSPKQ 230
Db 212 KQPEVPEEKAVEB-----TPKPEDKIKGIGTKEPVVK-SELNMQI-DKASSVSPD 261
QY 231 HQSALOSTDDEVVLVGPVLPQSSRLFTLKIRADLVRLPVMSF--PLQNVVD---- 284
Db 262 YSTASYAL-----GPVLETAKGYA-----SEPVKQPFVNSETNKLTALDALNV 307
QY 285 ---HMANHLGVSPRILLFGESELSPTAPSTLKLGVADIDCVVLASSEATETSQEL 341
Db 308 DKTELNTIADAKTKVEHSDRSWMLQTEVYTAKEVYAAANTD---AKQSEVNEAVEKL 363
QY 342 RLRYQKEKHQMLEISLSPDSPKLKVLMS 369
Db 364 TATIE-----KLVELS---EKPLITLS 383

RESULT 11
US-09-758-440-512
; Sequence 512, Application us/09758440
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PM015
; CURRENT APPLICATION NUMBER: US/09/758,440
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 832
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 512
; LENGTH: 227
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-758-440-512

Query Match 5.6%; Score 118; DB 5; Length 227;
Best Local Similarity 25.3%; Pred. No. 1.5;
Matches 62; Conservative 40; Mismatches 91; Indels 52; Gaps 11;

QY 1 MAEPLRGGRSGRGARRARAGRCPRARQSPARLIPDTVLVDLVSDS--DEVLEV 58
Db 8 VAVPLAGGQEGSPGG-GRKSRGT---TWYKKRGRVIDSDTDESGSDENLDQELLST 62
QY 59 A-----DPEVAVARLPYPAKPEODSDSEGAAGCAPARTLVRRRRRLDLDBGEA 111
Db 63 AKRRRSSEKEPEPVSO-PAASSDSETSDDDEWT---FGSNKKKKRKKRKEKKG-- 115
QY 112 PVVYVYSGKVOSSLNLIPDNSSLKICPSEPEDEADLTNSGSSPSEDALPSGSPWRK 171
Db 116 -TMKQANKTKASSGSSDKD--SSAESSAPEREGEVSDSDNSSSSSSDSDS----- 162
QY 172 RKCKEKKKEEPPD-----QDISPLPQPSRRKSRKHTALOKLREVNKRLOD 221
Db 163 -----SSEDEPHDYGEDLMGDEEDRARLEQMTKEKREQLEFNRIK-REVLKRFE 214
QY 222 LRSC 226

Db 215 IKKKL 219
RESULT 12
US-09-882-636-2
; Sequence 2, Application US/09882636
; GENERAL INFORMATION:
; APPLICANT: Botstein, David
; APPLICANT: Goddard, Audrey
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth
; APPLICANT: Lawrence, David, A
; APPLICANT: Roy, Margaret, Ann
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TREATMENT OF TUMOR
; FILE REFERENCE: P2509R1C1
; CURRENT APPLICATION NUMBER: US/09/882,636
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: 60/113,296
; PRIOR FILING DATE: December 22, 1998
; PRIOR APPLICATION NUMBER: 60/112,850
; PRIOR FILING DATE: December 16, 1998
; PRIOR APPLICATION NUMBER: 60/107,783
; PRIOR FILING DATE: November 10, 1998
; PRIOR APPLICATION NUMBER: 60/088,742
; PRIOR FILING DATE: June 10, 1998
; PRIOR APPLICATION NUMBER: 60/086,414
; PRIOR FILING DATE: May 22, 1998
; PRIOR APPLICATION NUMBER: 60/083,500
; PRIOR FILING DATE: April 29, 1998
; PRIOR APPLICATION NUMBER: 60/082,767
; PRIOR FILING DATE: April 23, 1998
; PRIOR APPLICATION NUMBER: 60/074,086
; PRIOR FILING DATE: February 9, 1998
; PRIOR APPLICATION NUMBER: 60/070,440
; PRIOR FILING DATE: January 5, 1998
; PRIOR APPLICATION NUMBER: 60/069,873
; PRIOR FILING DATE: December 17, 1997
; PRIOR APPLICATION NUMBER: 60/069,702
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069,694
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069,696
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/066,772
; PRIOR FILING DATE: November 24, 1997
; PRIOR APPLICATION NUMBER: 60/032,705
; PRIOR FILING DATE: December 12, 1996
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: December 1, 2000
; PRIOR APPLICATION NUMBER: 09/709,238
; PRIOR FILING DATE: November 8, 2000
; PRIOR APPLICATION NUMBER: 09/664,610
; PRIOR FILING DATE: September 18, 2000
; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: September 18, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/23522
; PRIOR FILING DATE: August 23, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/14941
; PRIOR FILING DATE: May 30, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/13705
; PRIOR FILING DATE: May 17, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/08439
; PRIOR FILING DATE: March 30, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/05841
; PRIOR FILING DATE: March 2, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/05004
; PRIOR FILING DATE: February 24, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: February 22, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/04341
; PRIOR FILING DATE: February 18, 2000


```

Db 1041 KRVSSEVCSSESVSYQVPAVADSEPTOKSFPPTKSDTLPSSHREIPLQD---DRLISLSPK 1097
;
;
;
QY 250 VLPOSSRLFTLKIRCRADLY-----RLPYRMSEPLQNVYDHNANHLGVSPN 295
;
;
;
Db 1098 ILPOQDNM-TAQDLAQREVMYSYKEKPOBELSLNKRKLKNSKSEAHITPSLFLKRETHS 1156
;
;
;
QY 296 RILLFESELSPT-----ATPSTLKL-GVADIIDCVVLASSSEATETSQE-LR 342
;
;
;
Db 1157 FIPLPFAEAKPKSTCELYSSQNEHAAPPSPNVIPGFQD-----RLSPSQSVLTPQODNLG 1211
;
;
;
QY 343 LRVOGKKKHOMLEISLSPDSLKY-----LMSHYEAMGLSGHKLSFFPGTKISG 393
;
;
;
Db 1212 LQKDLQREVLHYSQKAEKILVQRTALQOQIOKHKE-----TLKDFKDSQISK 1263
;
;
;
QY 394 KELPADLGESGDLIEVW 411
;
;
;
Db 1264 PTVENDLKTQKMGOLRDM 1281
;
;
;

```

```

RESULT 14
PCT-US01-14827-8085
;
; Sequence 8085, Application PC/TUS0114827
;
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-104
; CURRENT APPLICATION NUMBER: PCT/US01/14827
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 09/577,408
; PRIOR FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 16102
; SOFTWARE: Custom
; SEQ ID NO 8085
; LENGTH: 122
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (62)..(114)
; OTHER INFORMATION: ubiquitin domain proteins domain identified by eMATRIX.
; OTHER INFORMATION: accession number BL00299, P-value=2.250e-25, raw score of 28.84
PCT-US01-14827-8085

```

```

Query Match 5.6%; Score 116.5; DB 1; Length 122;
Best Local Similarity 28.4%; Pred. No. 0.79; 42; Mismatches 15; Gaps 2;
Matches 31; Conservative 21; Indels 15; Gaps 2;

```

```

QY 303 ESELPTATPTSTLKGAVADIIDCVVLASSSEATETSQELRLVQGEKHKOMLEISLSPDS 362
;
;
;
Db 23 EAARSPTGSGTSTREGV-----KTENNHNINKVAGOD-GSVVQFKIKRHT 67
;
;
;
QY 363 PLKVLMSHYEAMGLSGHKLSFFDGTGLSGKELPADLGESGDLIEVW 411
;
;
;
Db 68 PLSKIMRAYCEROGLSMKQIRFRFDGQPINETDPPAQLEMEDEDTIDVF 116
;
;
;

```

```

RESULT 15
US-09-758-438-749
;
; Sequence 749, Application US/09758438
;
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PM016
; CURRENT APPLICATION NUMBER: US/09/758,438
; CURRENT FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 1330
; SOFTWARE: PatentIn Ver. 2.0

```

```

;
; SEQ ID NO 749
;
; LENGTH: 458
;
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (110)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (124)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (418)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-758-438-749

```

```

Query Match 5.5%; Score 115.5; DB 5; Length 458;
Best Local Similarity 25.6%; Pred. No. 6;
Matches 63; Conservative 24; Mismatches 84; Indels 75; Gaps 11;

```

```

QY 34 SPARLIPPTVLVDLYSDSEVLEVADPEVPVAPRLPAPAKPEDSDSDSGAAGPAGA 93
;
;
;
Db 237 SSELSPDAVEKAGKSSNQ-----SISFVLDAVPTPTPERER-----SSASAPPMKDG 286
;
;
;
QY 94 PRTIVRRRRRLDPGEAPVYVYSGKVQSSINLIPDNSSL-----LKLCPSEPD 144
;
;
;
Db 287 PRTISRER-----SGSSPGLRDG-----TPSRHSLSGSSPGMKDIPRTSRGRS 333
;
;
;
QY 145 EADLTNCGSSPSEDDALP-----SGSPWRKKLRKKC--EKEKKMEFPDQ----- 188
;
;
;
Db 334 ECD-----SSP-EFKALPQTPRPRSRSPSPSELNNKCLTPQREKSGSSSVQKTVARTP 387
;
;
;
QY 189 -----DISPLPOPSSR-----NKSRRKHTALQKLRVYNNKRLQDRLSCL 227
;
;
;
Db 388 LGQSRSGSSQELDVKPSASPQENSESSXSQSRKAKTPTPLRQKRSKSSPEVDSKRLS 447
;
;
;
QY 228 PKOHQS 233
;
;
;
Db 448 PRSRRS 453
;
;
;

```

```

Search completed: September 11, 2001, 08:53:51
Job time: 325 sec

```

```

Db 121 VOSSLNLIIPNSSLKLCPEPEDEADLTNSGSSPSBEDDALPGSPWRKRLKCKEKEEK 180
QY 181 KMEFPPODISLPPOSSRNKSRKHTALOKLREVNKRLDOLRSCLSPKOHOSPALOSTD 240
    |||||
Db 181 KMEFPPODISLPPOSSRNKSRKHTALOKLREVNKRLDOLRSCLSPKOHOSPALOSTD 240
QY 241 DEVVLEGPVLPQSSRLFTLKIRCRADLVRLPVMSEPLQNVYDVMANHLGVSPNRILL 300
    |||||
Db 241 DEVVLEGPVLPQSSRLFTLKIRCRADLVRLPVMSEPLQNVYDVMANHLGVSPNRILL 300
QY 301 FGSELSPTATPSTLKGVADIDICVYLASSEETETSOELRLVQGEKHOMLEISLP 360
    |||||
Db 301 FGSELSPTATPSTLKGVADIDICVYLASSEETETSOELRLVQGEKHOMLEISLP 360
QY 361 DSPKLVMSHYEAMGLSGHKLSEFPDGTKLSGKELPADLGLSGDILEVWG 412
    |||||
Db 361 DSPKLVMSHYEAMGLSGHKLSEFPDGTKLSGKELPADLGLSGDILEVWG 412

```

```

RESULT 2
Q9GLZ9 PRELIMINARY; PTG: 408 AA.
AC Q9GLZ9:
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 44.6 KDA PROTEIN.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopitheinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN PARTIAL LOBE;
RA Otsuka N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.;
RT "Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries."
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB050511; BAB17279.1; -.
KW Hypothetical protein.
SQ
SEQUENCE 408 AA: 44580 MW: 5E06DD066AC2F24 CRC64:

```

```

Query Match 65.8%; Score 1382; DB 6; Length 408;
Best Local Similarity 70.1%; Pred. No. 2.4e-91;
Matches 293; Conservative 31; Mismatches 78; Indels 16; Gaps 5;

```

```

QY 1 MAEPLRGPRSRGGRGARRARGRCPRARQSPARLIPDTVLVDVSDSEEVLEV-- 58
    |||||
Db 1 MAEPLRGPRSRGGRGARRARGRCPRARQSPARLIPDTVLVDVSDSEEVLEV-- 58
QY 59 -----ADVEVAVARLPAPAKPEQSDSDSEGAAGPAGAPRTVLRRRRRLLDGEAPV 114
    |||||
Db 53 ARCADEVAVARLPAPAKPEQSDSDSEGAAGPAGAPRTVLRRRRRLLDGEAPV 112
    |||||
QY 115 PYVSGKVOSSLNLIIPNSSLKLCPEPEDEADLTNSGSSPSBEDDALPGSPWRKRLK 174
    |||||
Db 113 PYVSGKVOSSLNLIIPNSSLKLCPEPEDEADLTNSGSSPSBEDDALPGSPWRKRLK 171
    |||||
QY 175 CEKEEKKMEFPPODISLPPOSSRNKSRKHTALOKLREVNKRLDOLRSCLSPKOHOS 234
    |||||
Db 172 -DKREKKTEILDNDNSPLSPSPRTSKRHTALOKLREVNKRLDOLRSCLSPKOHOS 230
    |||||
QY 235 ALQSTDEVVLEGPVLPQSSRLFTLKIRCRADLVRLPVMSEPLQNVYDVMANHLGVSP 294
    |||||
Db 231 EQDQOEQFLKQHEFTLASSGAAAGAP--VQYVERSESISISRLFLQVOTPLREP-- 290
    |||||
QY 295 NRILLGESELSPTATPSTLKGVADIDICVYLASSEETETSOELRLVQGEKHOML 354
    |||||
Db 291 SRILLGESELSPTATPSTLKGVADIDICVYLASSEETETSOELRLVQGEKHOML 350
    |||||

```

```

QY 355 EISLSPDPLKVLMSHYEAMGLSGHKLSEFPDGTKLSGKELPADLGLSGDILEVWG 412
    |||||
Db 351 EISLSPDPLKVLMSHYEAMGLSGHKLSEFPDGTKLSGKELPADLGLSGDILEVWG 408
    |||||

```

```

RESULT 3
O60336 PRELIMINARY; PTG: 1217 AA.
AC O60336:
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE KIAA0596 PROTEIN (FRAGMENT).
GN KIAA0596.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE=98290545; Pubmed=9628581;
RA Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nomura N.,
RA Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro."
RL DNA Res. 5:31-39(1998).
DR EMBL: AB011168; BA25522.1; -.
DR InterPro: IPR001680; .
DR Pfam: PF00400; WD40; 7.
DR PRINTS: PR00320; GPROTEINBRPT.
DR SMART: SM00320; WD40; 1.
KW Repeat; WD repeat.
FT NON TER 1
SQ
SEQUENCE 1217 AA: 131097 MW: 0BC4E4C66722BEF5 CRC64:

```

```

Query Match 7.4%; Score 154.5; DB 4; Length 1217;
Best Local Similarity 23.5%; Pred. No. 0.0082;
Matches 105; Conservative 48; Mismatches 164; Indels 129; Gaps 21;

```

```

QY 6 RGRGPRSRGGRGARRARGRCPRARQSPARLIPDTVLVDVSDSEEVLEVADPYEVP 65
    |||||
Db 455 RGRGPRSRGGRGARRARGRCPRARQSPARLIPDTVLVDVSDSEEVLEVADPYEVP 65
QY 66 -----VARLPAPAKPEQSDSDSEGAAGPAGAPRTVLRRRRRLLDGEAPV 106
    |||||
Db 507 ALPVLAKSTKALASVSPALPNSLSHWEMSRQESVGFDPAPANPQP---RRGRWV 563
    |||||
QY 107 DPGEAPVAVPYVSGKVOSSLNLIIPNSSLKLCPE--EP-EDEADLTNSGSSPSBEDDALPS 163
    |||||
Db 564 QPG-----VELSVSMIDLRLQLETLAPSDQPSDLSALIPSGRKHQGEALLET 612
    |||||
QY 164 G-----SPWRKRLKRC--EKEE-----KMEFPPODISLPLOPS---SRNK 201
    |||||
Db 613 SITSQNEKPPRPOASQPCSPHILRLLSOEGVFAQDLLEAPLEDTGIVYDEPSNPTMDT 672
    |||||
QY 202 SRKHTALOK-----LREVNKRLQD-----LRSCLSPPKOHOSPALQSDDEEVVL 245
    |||||
Db 673 SEFOVQAPAKKTGLRGVPGSRSSSEKSPDACSVDYSSCSLSPFH--PLEDESSTEPS 730
    |||||
QY 246 VEGFVLPQSSRLFTLKIRCRADLVRLPVMSEPLQNVYDVMANHLGVSPNRILLFGSE 305
    |||||
Db 731 VDG-----ISSDLEFPADGDEEBEGGNGP-----YGLQ 762
    |||||
QY 306 LSPATP-----STLKGVADIDICVYLASSEETETSOELRLVQGEKHOMLE 355
    |||||
Db 763 GSP-QTPDQOEFLKQHEFTLASSGAAAGAP--VQYVERSESISISRLFLQVOTPLREP-- 818
    |||||
QY 356 ISLSPDPLKVLMSHYEAMGLSGHKLSEFPDGTKLSGKELPADLGLSGDILEVWG 412
    |||||
Db 819 ---SPSSSLALMSRPAQVPOASGEQ 841
    |||||

```


RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu L.B., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahler C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Mostneff A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wastarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Zibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003841; AAF59275.1;
 RL Flybase: FBgn003156; CG12165.
 SQ SEQUENCE 755 AA; 83536 MW; 5DA71535FE3EBCD1 CRC64;

Query Match 6.1%; Score 128.5; DB 5; Length 755;
 Best Local Similarity 24.1%; Pred. No. 0.33; Indels 73; Gaps 16;

Matches 77; Conservative 48; Mismatches 122; Indels 73; Gaps 16;

QY 71 APAKPEODSDSEGAEGAPART--LVRRRRRL-----DPEAEVVPVYSGKQ 122
 DB 41 AEAKPR-----DSAPAKSOENSSVTPQOTKKRRKRLTSLAEQNEAEADATANNNTSR 95
 QY 123 SSLNLIPDNSSLLKLCPEPEDEADLTNCGSSP--SEDDALPGSGPWR--KTLKKCEKEEK 180
 DB 96 QSTRV--SNSQLTIAEDENHSTASLMPPPVPSADTTGSGRPQRAKL-----KTEK 148
 QY 181 KMEFPDQDISLPQPSRRKSK--HTEAQLKLR--EVNKLQDLRSC-----LSP 228
 DB 149 LKE-----PSLNKKMRPSSSESLVKYKVESEQVSOFNSTSAOLEENKLA 197
 QY 229 KQHSPALOSTDDEVVLVEGPVLPOSSRLFTLKIRCRADLVRLVPRKSEPLQNVDMAN 288
 DB 198 PELEPRAETAEQOKPEEASVTEEDVTTTLTKYKAKR--KLTSEAVPPLTNVTS--TAN 254
 QY 289 HLGVPENRLLLGESELSPTATPSTLKLGVADIIDCVLSSSEATVSOELRLRVQK 348
 DB 255 VTVYS-----SVTTEAAR-----DDTVASNTTTSREVSKKVKKKKKV 293
 QY 349 EKHQMLEI-----SLSPDPL 364
 DB 294 ESHRPTEKVERPSDLDRSSPV 313

RESULT 13
 091995
 ID 091995 PRELIMINARY: PRT; 931 AA.
 AC 091995;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DE 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE CALPRSTATIN.
 GN CALP3.
 OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodidae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Maricci S., Rossi C., Nardi I.,
 RT "Identification of different forms of calpastatin mRNA co-expressed in
 RT the notochord of *Xenopus* embryos."
 RL Submitted (FE8-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: A271209; CAB71173.2;
 DR InterPro: IPR001259;
 DR Pfam: PF00748; Calpain_inhib. 5;
 SQ SEQUENCE 931 AA; 96825 MW; D71A35226255299E CRC64;

Query Match 6.1%; Score 128.5; DB 13; Length 931;
 Best Local Similarity 21.8%; Pred. No. 0.43; Indels 87; Gaps 18;

Matches 83; Conservative 60; Mismatches 150; Indels 87; Gaps 18;

QY 67 ARLPAPA--KEODSDSEGAEGAPARTLVRRRRRLDPEAP-----V 113
 DB 32 SKTPAPSGKTMMNPVDTPKGTATGAPAAATK-----PTAASALAKTTAAPS 79
 QY 114 VVYSGKVGSSNLIPDNSSLLKLCPEPEDEADLTNCGSSPSEDDALPGSGPWRKLR 173
 DB 80 VPAATKPTAA-----PSTAPVAVTKPTEIKRPSA-----GKSPKQDTRK--TPGKQKPS 128
 QY 174 KCEKEKKMEFPDQDISLPQ--PSSRNK-----SRKTEALQKREVNKR 218
 DB 129 K-SKEEKSDKATAPVOP--PKVTPSSAGKSAVATGFVAAGTAATVAVTQKODES 186
 QY 219 LQDLSCLSPK--OHQSPALOSTD--DEVVLVEGPVLPOSSRLFTLKIRCRADLVRLV 273
 DB 187 KEKKAVIGDKVSSAASPATGATFALDELDTIG-----SPADIPSPK 231
 QY 274 RMSEPLQ--VDMANHLG-----VSPNRLLLGESELSPTATPSTLKLGVADITDCV 326
 DB 232 FTGPEIOTDFTVTSKYVEELKRDHTIPPYRKLIDGKGEKMAPPTPLVVAESMDDDL 291
 QY 327 -VLASSEATETSELRLRVQK--EKHQMELISLSPDSPLKYLMSHYEAMGLSGHK- 381
 DB 292 AALSSGKSSQTCPEVKKPBLEEKLEPKSAATVATQDLOKVKVATAETSKASTOKK 351
 QY 382 --LSFPDGTGLSGELPAD 399
 DB 352 EVKATAEASKQSTQALPDS 371

RESULT 14
 09VN47
 ID 09VN47 PRELIMINARY: PRT; 1167 AA.
 AC 09VN47;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DE 01-JUN-2000 (Tremblrel. 14, Last annotation update)
 DE CG1054 PROTEIN.
 GN CG1054.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyoptera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abtill J.F., Agbayanti A., An H.-J., Andrews-Franknoch C., Baldwin D.,
RA Ballew R.M., Basu A., Bakendale B., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borckova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Butkiss K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferltera S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Gang N.S., Gelbart W.M., Glasser K.,
RA Glodet A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,
RA Jaitai M., Kalush F., Karen G.H., Ke Z., Kemnitz J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Krausz S., Kulp D., Lai Z.,
RA Lako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulyov G., Mishina N.V., Moberly C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclel J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Welnsloek G.M., Weissbach J.,
RA Williams S.M., Wodege T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zhang L.,
RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003605; AA952104.1; -
DR Flybase; FBgn0037267; CG1054.
SQ SEQUENCE 1167 AA: 126283 MW: 756CD221FBA8229D6 CRC64

Query Match	6.1%;	Score 127.5;	DB 5;	Length 1167;
Best Local Similarity	22.1%;	Pred. No. 0.67;		
Matches 86; Conservative	50;	Mismatches 122;	Indels 131;	Gaps 19;

```

0Y      PPTV--LVYDLYSDSDEEVL--EVADPVEVPARLP--APAKP-----EGDSDDSDS   83
        |::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      PPTVALNCFNSOSOSKSLTSPINSPPLPSGRILPQITPPASPPPKRVOLDDEVDRSSGD 703
        |::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y      EGAAGPAGA---PRTLVRRRRLRD-----PGEARVVVVYSKGKVQSSL 125
        |::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      YAPVVGQTALYHPPPSKTNANQOTLDPLAANDSLCLHTSPSELTSLACS----- 756
        |::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y      NLIPDNLSLLKCPSEPDEADLTGSGSGSPEDDALPGSPWRKLRLRK-----CEKE 178
        |::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      ---PNESSLEFLKASSRP-----TSPVDVA-----SCQFLNREAPKDIQLAEED 798
        |::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y      EKMEEFPPDO-----DISPLPOGSS-----RNRKRKTEALOKLREVKNRLQ    220
        |::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      RRRMDQFAVEEMPCKIDIVTSPKPQNTFGQYATEDLVKKCDGKNISAVNDL-DVPSQE 857
        |::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y      DIRSLSPKROHOSPALQSTDEVVYLVEGPLYPOSSRLFLLIKRCRADVLRLPYRMSEPIQ 280
        |::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      DIALCAGLLDLCLAPAQOEICDSIDVRNPVPATGFQ-----DISEEVKKCE--- 904
        |::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y      NVYDMAHHGLVSPKRIILLFGESELSTATPILKLGADVADIDCVYLASSSEVTETSOE 340
        |::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      -KENKNY-VITNR-----PSSSIGS-----VRAASHKALMDTCE 936
        |::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y      LRLRVQGEKHOMLEISL-----SPDSL 364
        |::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      OTIAEBKRASSOWMKTSLPESEVKRPSTPL 965
        |::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 15
Q9W0C9
ID Q9W0C9 PRELIMINARY; PRT; 1408 AA.

AC 09M0C9; 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 13, Last annotation update)
DE G212105 PROTEIN.
GN G212105.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Ephemeroptera; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;

```

RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=BERKELEY;
RX      MEDLINE=20196006; PubMed=10731132;

```

RA Adams M.D., Celniker S.E., Hol R.A., Evans C.A., Goezney J.D.,
RA Amaniatis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abtl J.F., Aygbanl A., An H.-J., Andrews-Piankoff C., Baldwin D.,
RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Besson K.Y., Benos P.V., Bernak B.P., Bhandari D., Bolshakov S.,
RA Borckova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Butts K.C., Busum D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahle C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Gloder A., Gong F., Gorell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibergwam C.,
RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kraivit S., Kuip D., Lal Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Milshina N.V., Moberry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Patel J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Pui V., Reese M.G.,
RA Rehner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Slden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spler E., Spradling A.C., Stepleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodgate T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.M., Rubin G.M., Venter J.C.,
RT "The genome sequence of *Drosophila melanogaster*."
RL Science 287:2185-2195(2000).
DR EMBL: AE003472; AAF47521.1; -
DR FlyBase: FBgn0035241; CG12105.
SQ FEEDBACK 1408 AA: 156836 MW: 7BAC33AC49574CD4 CRC64;

every Match	6.1%;	Score 127.5;	DB 5;	Length 1408;
1st Local Similarity	20.0%;	Pred. No. 0.84;		
Matches 94;	Conservative 50;	Mismatches 142;	Indels 185;	Gaps 20

QY	15	GGGARRRANGRCRRPARRQSPARLIPPTVY	-----DLVDS	-----DEEYLEVAD	60
Db	1002	GGGRT	-----RASLKLAEIVGETTVLVVAATAA	ODPLPESKEEIIIODEALVAT	1051
			-----AKPEODSDSDEGAEGACAP	RTLYRRRRRLDPG	109
QY	61	PVEVVPARLPAP	-----		
Db	1052	POQVAIGADIPQIAATLVHEESSPEQOSS	SEATKAGEOPPGSGIATMPELV	---	1108
			-----EPE	-----DEADLTN	150
QY	110	EAPVVPYSGKVOSSLNLPIDNSSLKLCTCS	-----		
Db	1109	AGPQVPVITVEQPHQAANKQOEIVLKKR	PTHLKLRKNSIDKEEKEKNDSSKLNI		1168

```
QY 151 -----SGSSPSEDDALP-----SGSPWRKKLRKKCEKEKME 183
Db 1169 LRDNKINELSSRINKLTPPKQAKPKKEAKKATTANEDSAKPEEATPAEETTAETPALE 1228
QY 184 EF-PDODISPLPQSSRNKSRKHTREALQKLEVNKRLODLRCLSPKOHOSPALOS---- 238
Db 1229 EVEPEVEPAPEPKPKRIKKV---IKRQKRRLSVGDTFFLOPEPEPEPAIPEVETI 1284
QY 239 -----TDDEVVLVEGPVLPOSSRLFTLIRCRADLVRLPVRMSEPLQNVVDHMANHLG 291
Db 1285 ERAIAYVTDE---EDP-----KAPAEPEPEV----- 1308
QY 292 VSPNRILLFGESELSPAPSTLKLGVADIIDCVLASSSEATE-----TSQ 339
Db 1309 -----NELKSCLVHREYKIG-----DIVLYAERYRKTQVRKKRGVLERITSI 1351
QY 340 ELRLVQGKEKHOMLEISLSPDSPLVLMSHYEAMGISGHKLSFFPDGTR 390
Db 1352 SYKLEIEGKE-----VPAHVSYIKKYTGRRKRF---GTK 1382
```

Search completed: September 11, 2001, 08:56:05
Job time: 429 sec

This Page Blank (uspo)

~~BEST AVAILABLE COPY~~

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 11, 2001, 08:48:26 ; Search time 19.29 Seconds
(without alignments)
1294.820 Million cell updates/sec

Title: US-09-192-611-2

Perfect score: 2099
Sequence: 1 MAEPLRGGRPSRSGRGARR.....GKELPADLGLESGLIEWVG 412

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_0601:*

1:	/SIDSI/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2:	/SIDSI/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3:	/SIDSI/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4:	/SIDSI/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5:	/SIDSI/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6:	/SIDSI/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7:	/SIDSI/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8:	/SIDSI/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9:	/SIDSI/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10:	/SIDSI/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11:	/SIDSI/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12:	/SIDSI/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13:	/SIDSI/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14:	/SIDSI/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15:	/SIDSI/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16:	/SIDSI/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17:	/SIDSI/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18:	/SIDSI/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19:	/SIDSI/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20:	/SIDSI/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21:	/SIDSI/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22:	/SIDSI/gcgdata/geneseq/geneseqp/AA2001.DAT:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2099	100.0	412	18	AAW34091
2	2099	100.0	412	20	AAV08331
3	1399.5	66.7	419	20	AAV08330
4	132	6.3	586	18	AAW10423
5	124.5	5.9	518	21	AAW53322
6	124.5	5.9	671	21	AAV9426
7	124.5	5.9	671	22	AAW66175
8	123	5.9	950	20	AAV33298
9	122.5	5.8	1780	19	AAW53863
10	122.5	5.8	1780	21	AAW15380
11	121	5.8	2971	21	AAW41231

12	121	5.8	2972	22	AAW50363
13	121	5.8	3118	22	AAW50362
14	120.5	5.7	565	19	AAW61247
15	120.5	5.7	1881	21	AAW44506
16	120	5.7	669	19	AAW37483
17	118.5	5.6	508	21	AAW77945
18	118.5	5.6	846	21	AAW71057
19	118	5.6	576	20	AAW49541
20	118	5.6	576	20	AAW49545
21	118	5.6	576	20	AAW49546
22	118	5.6	576	20	AAW06477
23	118	5.6	576	21	AAW93684
24	118	5.6	576	21	AAW51932
25	118	5.6	576	21	AAW51939
26	117	5.6	979	14	AAW34783
27	115.5	5.5	1341	21	AAW85657
28	115	5.5	576	20	AAW49547
29	113.5	5.4	661	19	AAW77048
30	113.5	5.4	661	21	AAW70078
31	113	5.4	519	21	AAW44247
32	113	5.4	783	20	AAW60344
33	112.5	5.4	95	20	AAW87985
34	112.5	5.4	610	12	AAW10923
35	111.5	5.3	244	21	AAW16198
36	111.5	5.3	2286	22	AAW65635
37	111	5.3	605	20	AAW31741
38	111	5.3	897	17	AAW92751
39	111	5.3	887	19	AAW47118
40	111	5.3	897	20	AAW94406
41	110.5	5.3	1519	21	AAW85660
42	110.5	5.3	2518	21	AAW40574
43	109.5	5.2	605	21	AAW57950
44	109.5	5.2	710	20	AAW80995
45	109.5	5.2	1142	21	AAW43876

ALIGNMENTS

RESULT 1					
AAW34091	ID	AAW34091 standard; Protein; 412 AA.			
XX	AC	AAW34091;			
XX	DT	18-MAY-1998 (first entry)			
XX	DE	Mouse NF-AT interacting protein 45.			
XX	KW	NF-AT Interacting Protein 45; NIP45; yeast two-hybrid assay; mouse;			
KW	Rel	homology domain; RHD; T cell; transcription factor; cancer;			
KW	interleukin-4; IL-4;	development; Th1; Th2; cytokine; allergy;			
KW	autoimmune disease; transplantation.				
XX	XX				
OS	Mus sp.				
XX	XX				
XX	Key	Location/Qualifiers			
FT	Region	6..37			
FT		/note="highly basic region"			
XX	PN	W09739721-A2.			
XX	XX				
PD	30-OCT-1997.				
XX	XX				
PF	23-APR-1997;	97WO-US06708.			
XX	XX				
XX	25-NOV-1996;	96US-0755592.			
PR	23-APR-1996;	96US-0636602.			
XX	PR	25-NOV-1996;	96US-0755584.		
XX	XX				
PA	(HARD)	HARVARD COLLEGE.			
XX	XX				
PI	Glincher LH; Ho I; Hodge MR;				

Human SRCAP. Homo
Human SRCAP. Homo
Streptococcus pneu
Streptococcus pneu
Mouse liver cancer
A. thaliana enviro
Human membrane tra
Human PRO201 (Nsp1)
Human Nsp1 mutant
Human Nsp1 mutant
Human tumour-assoc
Amino acid sequenc
Human PRO201 prote
Human PRO201 prote
Human CENP-C anti
Human Actinus L pro
Human Nsp1 mutant
Human striated mus
Human striated mus
Human cell signal
Human normal blad
Ubiquitin-like dom
Maize protein enco
Arabidopsis thalia
Novel protein kina
Human podocalyxin-
Murine EGF recepto
Murine eps15 prote
Human eps15 prote
Human tyrosine kin
Human ORFX ORF338
Human transmembran
Human guanine nucl
Amino acid sequenc

```

XX      WPI: 1997-535556/49.
DR      N-PSDB; AAT93045.
XX
XX      Production of cytokine(s) associated with Th2-type helper T cells -
PT      particularly for controlling development of Th1 and Th2 cells for
PT      treatment of allergy, autoimmune disease etc.
XX
XX      Claim 31: Fig 11; 151pp; English.
XX
CC      This is the amino acid sequence of the mouse NF-AT interacting Protein 45
CC      (NIP45). The gene sequence was isolated by using a yeast two-hybrid
CC      detection system for proteins that interact with the NF-AT Rel homology
CC      domain (RHD). The assay used, as a "bait", a 900 bp fragment of the
CC      murine NF-ATp encoding the region spanning amino acids 228-250. NF-AT is
CC      a multisubunit transcription complex containing a cyclosporin A sensitive
CC      cytoplasmic phosphoprotein and an inducible component of the AP-1 family
CC      of transcription factors. The screen was carried out on a cDNA library
CC      prepared from the murine T cell line D10. One class of proteins, NIP45
CC      designated NIP45, bound the NF-AT-RHD region with high affinity. NIP45
CC      can be used in a claimed method to inhibit or stimulate production of
CC      NF-AT family protein, particularly interleukin-4 (IL-4) and the treated
CC      cells may be administered to control development of Th1 or Th2 cells by
CC      modulating the production of a T-helper type 2 associated cytokine.
CC      Especially the method is useful in the inhibition of Th2 in allergy,
CC      cancer or infections, and promotion of Th2 in autoimmune disease and
CC      transplantation.
XX
XX      Sequence 412 AA:
SQ

```

```

Query Match      100.0%; Score 2099; DB 18; Length 412;
Best Local Similarity 100.0%; Pred. No. 7.4e-175;
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 MAEPLRGGRGRSGRGARRARGRCPRAROSPARLIPDYLVLDVSDSEVLEVAD 60
DB      1 MAEPLRGGRGRSGRGARRARGRCPRAROSPARLIPDYLVLDVSDSEVLEVAD 60
QY      61 PVEVPVRLPAPAKPEQSDSDSEGAEGPAGAPRTLVRRRRRLDPGEAPVVPVYSGK 120
DB      61 PVEVPVRLPAPAKPEQSDSDSEGAEGPAGAPRTLVRRRRRLDPGEAPVVPVYSGK 120
QY      121 VQSSSLNLIIPNSSLKLCPEPEDEADLTNSGSSPSDDALPSGSPRRKLRKKCEEEK 180
DB      121 VQSSSLNLIIPNSSLKLCPEPEDEADLTNSGSSPSDDALPSGSPRRKLRKKCEEEK 180
QY      121 VQSSSLNLIIPNSSLKLCPEPEDEADLTNSGSSPSDDALPSGSPRRKLRKKCEEEK 180
DB      121 VQSSSLNLIIPNSSLKLCPEPEDEADLTNSGSSPSDDALPSGSPRRKLRKKCEEEK 180
QY      181 KMEFPFQDLSPIPPQSSSRKSKKHTPEALOKLREVNKRLODLRSCSPKOHQSPALOSTD 240
DB      181 KMEFPFQDLSPIPPQSSSRKSKKHTPEALOKLREVNKRLODLRSCSPKOHQSPALOSTD 240
QY      241 DEVVVEGPPVLPQSSRLFTLKIRCRADLVRLPYMSEPLQNVVDHMANHLGVSPNRITLL 300
DB      241 DEVVVEGPPVLPQSSRLFTLKIRCRADLVRLPYMSEPLQNVVDHMANHLGVSPNRITLL 300
QY      301 FGESELSPTATPSTLKLGVADIIDCVVLASSSEATETSQELRLRVGKERKHQMLEISLSP 360
DB      301 FGESELSPTATPSTLKLGVADIIDCVVLASSSEATETSQELRLRVGKERKHQMLEISLSP 360
QY      361 DSPPLKVLMSHYEAMGLSGHKLSFFPDGTLKSGKELPADGLGSGDILEWVG 412
DB      361 DSPPLKVLMSHYEAMGLSGHKLSFFPDGTLKSGKELPADGLGSGDILEWVG 412

```

```

RESULT 2
ID      AAAY08331 standard; Protein; 412 AA.
XX
XX      AAAY08331;
AC
XX      16-jUL-1999 (first entry)
XX
XX      Mouse NIP45 protein.
DE

```

```

XX      NIP45; mouse; transcription transactivator; IL-4; interleukin-4;
KW      NF-AT interacting protein-45; modulator; treatment; inflammation;
KW      autoimmune disease; HTP screening; drug testing; allergic therapy;
KW      T-cell dependent response; cytokine; diagnostic; immunosuppressant.
XX
XX      Mus musculus.
XX
XX      MO9921993-A1.
XX
XX      06-MAY-1999.
XX
XX      21-OCT-1998; 98WO-GB03141.
XX
XX      24-OCT-1997; 97GB-0022388.
XX
XX      (ZENE) ZENECA LTD.
XX
XX      Liu D, Zhao J, Zhou H;
XX
XX      WPI: 1999-312964/26.
XX
XX      Polynucleotides encoding human NF-AT interacting protein (NIP45)
PT      disclosure; Fig 4; 84pp; English.
XX
XX      This invention describes a novel human NF-AT interacting protein-45
XX      NIP-45. Human NIP45 is a transcriptional trans-activation factor of the
XX      interleukin 4 (IL-4) gene. Modulators of human NIP45 and IL-4 can be used
XX      to treat conditions mediated by NIP45 or IL-4, e.g. inflammation or
XX      autoimmune disease. NIP45 is a good candidate target for HTP screening
XX      and/or testing system for drugs that will alleviate T-cell dependent
XX      autoimmune and allergic responses, and for cytokine-based therapies of
XX      chronic disease. Expression of NIP45 can be inhibited, and IL-4
XX      expression modulated, by administering an antisense molecule. Antibodies
XX      against NIP45 and primers can be used in diagnostic assays. Discovery of
XX      a human NIP45 is advantageous in that it provides the ability to control
XX      IL-4 transcription, which is of importance for anti-inflammation and
XX      immunosuppressant drug development.
XX
XX      Sequence 412 AA:
SQ

```

```

Query Match      100.0%; Score 2099; DB 20; Length 412;
Best Local Similarity 100.0%; Pred. No. 7.4e-175;
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 MAEPLRGGRGRSGRGARRARGRCPRAROSPARLIPDYLVLDVSDSEVLEVAD 60
DB      1 MAEPLRGGRGRSGRGARRARGRCPRAROSPARLIPDYLVLDVSDSEVLEVAD 60
QY      61 PVEVPVRLPAPAKPEQSDSDSEGAEGPAGAPRTLVRRRRRLDPGEAPVVPVYSGK 120
DB      61 PVEVPVRLPAPAKPEQSDSDSEGAEGPAGAPRTLVRRRRRLDPGEAPVVPVYSGK 120
QY      121 VQSSSLNLIIPNSSLKLCPEPEDEADLTNSGSSPSDDALPSGSPRRKLRKKCEEEK 180
DB      121 VQSSSLNLIIPNSSLKLCPEPEDEADLTNSGSSPSDDALPSGSPRRKLRKKCEEEK 180
QY      121 VQSSSLNLIIPNSSLKLCPEPEDEADLTNSGSSPSDDALPSGSPRRKLRKKCEEEK 180
DB      121 VQSSSLNLIIPNSSLKLCPEPEDEADLTNSGSSPSDDALPSGSPRRKLRKKCEEEK 180
QY      181 KMEFPFQDLSPIPPQSSSRKSKKHTPEALOKLREVNKRLODLRSCSPKOHQSPALOSTD 240
DB      181 KMEFPFQDLSPIPPQSSSRKSKKHTPEALOKLREVNKRLODLRSCSPKOHQSPALOSTD 240
QY      241 DEVVVEGPPVLPQSSRLFTLKIRCRADLVRLPYMSEPLQNVVDHMANHLGVSPNRITLL 300
DB      241 DEVVVEGPPVLPQSSRLFTLKIRCRADLVRLPYMSEPLQNVVDHMANHLGVSPNRITLL 300
QY      301 FGESELSPTATPSTLKLGVADIIDCVVLASSSEATETSQELRLRVGKERKHQMLEISLSP 360
DB      301 FGESELSPTATPSTLKLGVADIIDCVVLASSSEATETSQELRLRVGKERKHQMLEISLSP 360
QY      361 DSPPLKVLMSHYEAMGLSGHKLSFFPDGTLKSGKELPADGLGSGDILEWVG 412
DB      361 DSPPLKVLMSHYEAMGLSGHKLSFFPDGTLKSGKELPADGLGSGDILEWVG 412

```



```
Dh 11 rarrprarrg99g9vsnsrths-gkcrirgralsapplflattttttmmyvastddds 69
QY 56 LEVADPEVVALPAPAKREQSDSDSEGAAGPACAPRTIVRRRRRLDDGCEAVVP 115
Db 70 lllktpdel-----dkysgsptlll-----tltdkhdirqpr 101
QY 116 VYSGK---VOSINLIPDN-----SSLKLCPEPEDEADLTNCGSPSEDALPS 163
Db 102 vhrqcyhlhldirpeelrpfqlllstrplq--geandeq--taptllgeeeetaas 157
QY 164 GSWRRKKLRKCKEKEKMEFPDODISPLPQSSRNKSKRHTEALOKLREVNRLQDLR 223
Db 158 hepekk-----ekgekke-----drddrer-----gllvsnedsdvr 197
QY 224 SCLSS--PKHQSPALOSTDDEVVLVEGVPVLPOSSRLFTLIRCRADIVRLPVRMSEPLON 281
Db 198 pafslfparpgchllrsvldq-qltmaivrlhnlfaltl-llplkrlplr-----rk 250
QY 282 VDHMANHLGVSPNRILLFGESELSPTAPSTPLKLGADIIDCVLASSSEATERTSQEL 341
Db 251 aahhtah-----dcalahlpeltfepdlmvtenaasvad---taestda-dltppl 301
QY 342 RLAVQCKEKHOMLEISLSPSPPLKVLMSHYEAMGUSG-----HKLS----- 383
Db 302 tvtvr-----halcwhrveg-glsqprgltsrtsarlsetaktlqp 342
QY 384 FFPDGRFKLSGKELPADLGLES 404
Db 343 svfgrleldpneapdltlss 363

RESULT 5
AAB53322
ID AAB53322 standard; Protein; 518 AA.
XX
AC AAB53322;
XX
DT 09-MAR-2001 (first entry)
XX
DE Human colon cancer antigen protein sequence SEQ ID NO:862.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW identification; cytosolic; cardioactive; neuroprotective; vulnery;
KW immunomodulatory; muscular; gynaecological; gastrointestinal;
KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
KW neuropathic disorder; immune system disorder; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; renal disorder;
KW infectious disease; cardiovascular disorder.
XX
XX Homo sapiens.
XX OS
XX PN MO200055351-A1.
XX PD 21-SEP-2000.
XX PF 08-MAR-2000; 2000MO-US05883.
XX PR 12-MAR-1999; 99US-0124270.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA; Ruben SM;
XX DR WPI; 2000-587534/55.
XX DR N-PSDB; AAC98079.
XX
XX Colon cancer associated gene sequences, referred to as colon cancer
XX PT antigens, useful for the treatment, prevention, and diagnosis of colon
XX disorders such as colon cancer -
XX
XX C1a4m 11; Page 1414-1416; 2104pp; English.
XX PS
XX AAC97991 t9 AAC98763 encode the human colon cancer associated proteins,
```

```
CC called human colon cancer antigens, given in AAB53324 to AAB54006. The
CC human colon cancer antigens can have cytotoxic, cardioactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC vulnery, nephrotropic, antiinfective and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins
CC may also be used to prevent diseases such as neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, wounds, renal disorders, infectious
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC AAB54007 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 518 AA;
XX
Query Match 5.9%; Score 124.5; DB 21; Length 518;
Best Local Similarity 23.1%; Pred. No. 0.02;
Matches 93; Conservative 41; Mismatches 143; Indels 125; Gaps 20;
QY 8 RGPSSRCGARGARAGRCPRAROSPAPRLIPDTYLVLDLV-----SDSD 52
Db 97 rapr-rgplggrkk-----kapsadsdkadsdgakpepvamarasasssssssdqd 151
QY 53 EBYLE-----VADPEVVALPAPAKP-----EDSDSD-----SEGAABGPACAPRT 96
Db 152 vsvkkprrgkpkpaekpikprrgkpkprrpsasssdsddevdrisewkrdrrel 211
QY 97 LVRRRR-----RRL-----LDPEGAPVVPYSGKVOSSLNLIIPDNSSLK 136
Db 212 earrirgeeelrrlregekekerrrrdrgeae-----rsgsgsgdelredepvk 266
QY 137 -----LCPSEPPD--EADLTNSGSSPEDDALPSSPWRKRLR----- 172
Db 267 kgrtkgrgrgppssdsdsepealeareakkaakpqsststeparkpqkkrvprkekqqa 326
QY 173 --KKCEKEKKMEFP--DODISPLPQSSRNKSRKHTKRALOKLR--EVNKRLO----DLRS 224
Db 327 kpykvertkrrsegfsmdrivvekkkepsvee-----lqklnseikfalkvdsdpvkr 379
QY 225 CLS-----PKHQSPALOSTDDEVVLVE-----GPVLPOSSRLFTLIRCRADIVR 270
Db 380 clnaleelgltgtsgllqntlvatllkllrrykankdkvmeakaeyt-----r 429
QY 271 LPVRMSEPLONVVDHMANHLGVSPNRILLFGESELSPTATP 312
Db 430 lksrvtlqpklaevqkv-nkagmekkkaeklaageelaageap 470

RESULT 6
AAY99426
ID AAY99426 standard; Protein; 671 AA.
XX
AC AAY99426;
XX
DT 08-AUG-2000 (first entry)
XX
DE Human PRO1604 (UNG785) amino acid sequence SEQ ID NO:308.
XX
KW Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
KW transmembrane; secretion; immunoadhesion; pharmaceutical; screening.
XX
XX Homo sapiens.
XX OS
XX PN MO200012708-A2.
XX PD 09-MAR-2000.
XX PF 01-SEP-1999; 99MO-US20111.
XX
```



```
PR 01-SEP-1998; 98US-0098716.
PR 01-SEP-1998; 98US-0098749.
PR 01-SEP-1998; 98US-0098750.
PR 02-SEP-1998; 98US-0098803.
PR 02-SEP-1998; 98US-0098821.
PR 02-SEP-1998; 98US-0098843.
PR 09-SEP-1998; 98US-0099536.
PR 09-SEP-1998; 98US-0099596.
PR 09-SEP-1998; 98US-0099598.
PR 09-SEP-1998; 98US-0099602.
PR 09-SEP-1998; 98US-0099642.
PR 10-SEP-1998; 98US-0099741.
PR 10-SEP-1998; 98US-0099754.
PR 10-SEP-1998; 98US-0099763.
PR 10-SEP-1998; 98US-0099792.
PR 10-SEP-1998; 98US-0099808.
PR 10-SEP-1998; 98US-0099812.
PR 10-SEP-1998; 98US-0099815.
PR 10-SEP-1998; 98US-0099816.
PR 15-SEP-1998; 98US-0100385.
PR 15-SEP-1998; 98US-0100388.
PR 15-SEP-1998; 98US-0100390.
PR 16-SEP-1998; 98US-0100584.
PR 16-SEP-1998; 98US-0100587.
PR 16-SEP-1998; 98US-0100661.
PR 16-SEP-1998; 98US-0100664.
PR 17-SEP-1998; 98US-0100683.
PR 17-SEP-1998; 98US-0100684.
PR 17-SEP-1998; 98US-0100710.
PR 17-SEP-1998; 98US-0100711.
PR 17-SEP-1998; 98US-0100919.
PR 17-SEP-1998; 98US-0100930.
PR 18-SEP-1998; 98US-0100848.
PR 18-SEP-1998; 98US-0100849.
PR 18-SEP-1998; 98US-0101014.
PR 18-SEP-1998; 98US-0101068.
PR 18-SEP-1998; 98US-0101071.
PR 22-SEP-1998; 98US-0101279.
PR 23-SEP-1998; 98US-0101471.
PR 23-SEP-1998; 98US-0101472.
PR 23-SEP-1998; 98US-0101474.
PR 23-SEP-1998; 98US-0101475.
PR 23-SEP-1998; 98US-0101476.
PR 23-SEP-1998; 98US-0101477.
PR 23-SEP-1998; 98US-0101479.
PR 24-SEP-1998; 98US-0101738.
PR 24-SEP-1998; 98US-0101741.
PR 24-SEP-1998; 98US-0101743.
PR 24-SEP-1998; 98US-0101915.
PR 24-SEP-1998; 98US-0101916.
PR 29-SEP-1998; 98US-0102207.
PR 29-SEP-1998; 98US-0102240.
PR 29-SEP-1998; 98US-0102307.
PR 29-SEP-1998; 98US-0102330.
PR 29-SEP-1998; 98US-0102331.
PR 30-SEP-1998; 98US-0102484.
PR 30-SEP-1998; 98US-0102487.
PR 30-SEP-1998; 98US-0102570.
PR 30-SEP-1998; 98US-0102571.
PR 01-OCT-1998; 98US-0102571.
PR 01-OCT-1998; 98US-0102684.
PR 01-OCT-1998; 98US-0102687.
PR 02-OCT-1998; 98US-0102965.
PR 06-OCT-1998; 98US-0103258.
PR 06-OCT-1998; 98US-0103449.
PR 07-OCT-1998; 98US-0103314.
PR 07-OCT-1998; 98US-0103315.
PR 07-OCT-1998; 98US-0103328.
PR 07-OCT-1998; 98US-0103395.
PR 07-OCT-1998; 98US-0103396.
PR 07-OCT-1998; 98US-0103401.
PR 08-OCT-1998; 98US-0103633.
PR 08-OCT-1998; 98US-0103678.

PR 08-OCT-1998; 98US-0103679.
PR 08-OCT-1998; 98US-0103711.
PR 14-OCT-1998; 98US-0104257.
PR 20-OCT-1998; 98US-0104987.
PR 20-OCT-1998; 98US-0105000.
PR 20-OCT-1998; 98US-0105002.
PR 21-OCT-1998; 98US-0105104.
PR 22-OCT-1998; 98US-0105169.
PR 22-OCT-1998; 98US-0105266.
PR 26-OCT-1998; 98US-0105693.
PR 26-OCT-1998; 98US-0105694.
PR 27-OCT-1998; 98US-0105807.
PR 27-OCT-1998; 98US-0105881.
PR 27-OCT-1998; 98US-0105882.
PR 27-OCT-1998; 98US-0106062.
PR 28-OCT-1998; 98US-0106023.
PR 28-OCT-1998; 98US-0106029.
PR 28-OCT-1998; 98US-0106030.
PR 28-OCT-1998; 98US-0106032.
PR 28-OCT-1998; 98US-0106033.
PR 28-OCT-1998; 98US-0106178.
PR 29-OCT-1998; 98US-0106248.
PR 29-OCT-1998; 98US-0106384.
PR 29-OCT-1998; 98US-0108500.
PR 30-OCT-1998; 98US-0106464.
PR 03-NOV-1998; 98US-0106856.
PR 03-NOV-1998; 98US-0106902.
PR 03-NOV-1998; 98US-0106905.
PR 03-NOV-1998; 98US-0106919.
PR 03-NOV-1998; 98US-0106932.
PR 03-NOV-1998; 98US-0106934.
PR 10-NOV-1998; 98US-0107783.
PR 17-NOV-1998; 98US-0108775.
PR 17-NOV-1998; 98US-0108779.
PR 17-NOV-1998; 98US-0108787.
PR 17-NOV-1998; 98US-0108788.
PR 17-NOV-1998; 98US-0108801.
PR 17-NOV-1998; 98US-0108802.
PR 17-NOV-1998; 98US-0108806.
PR 17-NOV-1998; 98US-0108867.
PR 17-NOV-1998; 98US-0108925.
PR 18-NOV-1998; 98US-0108848.
PR 18-NOV-1998; 98US-0108849.
PR 18-NOV-1998; 98US-0108850.
PR 18-NOV-1998; 98US-0108851.
PR 18-NOV-1998; 98US-0108852.
PR 18-NOV-1998; 98US-0108858.
PR 18-NOV-1998; 98US-0108904.

PA (GETH ) GENENTECH INC.
XX
XX Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WT:
XX WPI: 2000-237871/20.
XX N-PSDB: AAA37108.
XX
XX New mammalian DNA sequences encoding transmembrane, receptor or
PT secreted PRO polypeptides, useful for screening of potential peptide or
PT small molecule inhibitors of the relevant receptor/ligand interactions
XX
XX Claim 12: Fig 174; 773bp; English.
XX
XX AAA37022 to AAA37144 encode the new isolated human transmembrane,
XX receptor or secreted PRO polypeptides given in AAy99340 to AAy99462. The
XX transmembrane and receptor PRO proteins can be used for screening of
XX potential peptide or small molecule inhibitors of the relevant
XX receptor/ligand interactions. The polypeptides and nucleotide sequences
XX encoding them have various industrial applications, including uses as
XX pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
XX PCR primers and hybridisation probes used in the isolation of the PRO
XX polypeptides from the present invention.
XX
```


QY	310	A	TPSRLTKIGVADIIDCVVLASSSEATFESQELRLR---	YOGKEHQMLEISLSPDPL	364			
Db	438	agvsaaeeelvgmd-----	aeppgaapekkelvklctvsgedpctgad--	lspdkv	487			
QY	365	-----	KVLMISHYEAMGLSGHKLSFFPDGT---	KLSGKELPADIG--	LESGDLI 408			
Db	488	lskppegvsevemlssqermkvqgsplklkflfstlgiklsgkkkqkkggdsgeht			547			
QY	409	EV 410	:	:				
Db	548	qv 549	:	:				
RESULT 10								
AAB15380								
ID	AAB15380 standard; Protein; 1780 AA.							
XX	AAB15380;							
AC								
XX	26-JAN-2001 (first entry)							
DT								
XX	Human gravin protein sequence.							
DE								
KW	Human; gravin; PKA RII binding site; myasthenia gravis;							
KM	kinase anchoring protein; CAMP dependent protein kinase.							
XX	Homo sapiens.							
OS								
XX								
FH	Key	Location/Qualifiers						
FT	Binding-site	265..556	/note="PKC binding site"					
FT	Blinding-site	1526..1582	/note="PKA RII binding site"					
FT	Region	1537..1563	/note="PKA anchoring site"					
FT								
XX	US6090929-A.							
PN								
XX	18-JUL-2000.							
PD								
XX	19-DEC-1997; 97US-0094570.							
PE								
XX	19-DEC-1996; 96US-0769309.							
PR								
XX	(UYOR-) UNIV OREGON HEALTH SCI.							
PA								
XX	Klauck TM, Scott JD, Nauert JB;							
PI								
XX	WPI: 2000-523763/47.							
DR	N-PSDB; AAA/4903.							
DR								
XX	Novel polynucleotides useful for detecting gravin in patients suffering							
PT	from Myasthenia gravis encodes CAMP-dependent protein kinase-binding							
PT	polypeptide and protein kinase C-binding polypeptide of gravin -							
XX								
XX	Claim 1; Column 35-45; 34pp: English.							
PS								
XX								
CC	The present sequence is the protein sequence of human gravin. Gravin is							
CC	an A-kinase anchoring protein (AKAP) which is involved in the							
CC	localisation of CAMP dependent protein kinase A (PKA) via interactions							
CC	between the RII binding region and the PKA regulatory subunit RII. Gravin							
CC	is also an antigen found in myasthenia gravis sufferers, and it is							
CC	thought that antibodies to it may be useful in modulating the binding of							
CC	PKA, and thus aid in the treatment of the disease. The gravin coding							
CC	sequence was isolated by first screening a human umbilical vein							
CC	endothelial cell cDNA library with serum from a myasthenia gravis							
CC	patient, and then searching a human heart cDNA library for sequences							
CC	resembling the isolated sequence. This was done because the first							
CC	sequence obtained was shown to be shorter than the full length cDNA.							
XX								
XX	Sequence 1780 AA;							

Query Match	5.8%;	Score 122.5;	DB 21;	Length 1760;
Best Local Similarity	22.0%;	Pred. No. 0.18;		
Matches	93;	Conservative	62;	Mismatches 152;
			Indels	115;
			Gaps	22;

QY	40	PDIV-LVDVLVSDSEEVLEVADPVEVYVRLPAK---	FEQSDSDSSEGAAGCAPAR	95
Db	192	pdlvylilivkkdegeagaagdhqdpislgaaaskesepkqstlekp	-----	240
QY	96	TLVRRRRRLLDPCGAPVYVSGKVOSSLNLIIDNSSLLKLCSPED--	EADLTNGS	153
Db	241	tlkregshaelisp	-----paesgq	281
QY	154	SPSEDDALPSCSP-----	WRKK--LRKCEKE--EKKMEFPDQDISPLPQ	196
Db	282	splsvtsetstgftfkfftgwagwrkktstfrpkedevasekkqepdek	-----vdl	336
QY	197	SSRNKSRKHTEALQKLRRVNNKRIODLRGCLSPK--- <th>QHOSPALQSTUDEVLVEGP---</th> <td>249</td>	QHOSPALQSTUDEVLVEGP---	249
Db	337	eedgvaevasekltasegahpq--paaeahaprIlaeekvelps--	eegvsqsgpreek	354
QY	250	VLPQSRSLFTLKIRCRADLVRLPRMSEPLQNVVDHMANHGLVSPNRILLFGSELSPT	309	
Db	335	paplaterfdekienhgeevaaehvstveerteq	-----ktevec	437
QY	310	A--TPSTKLGVADIIDCVLASSEATETSQELRLR--- <th>VQGEKHOMLEISLSPDSL</th> <td>354</td>	VQGEKHOMLEISLSPDSL	354
Db	438	agvsphaelvgmd-----	aepqaaepakelvkiketcvsgdpcqgad--lspek	487
QY	355	----- <th>VLMSHYEAMGLSGHKLSPEFDG---KLSGKRLPDLG--LESGDLI</th> <td>408</td>	VLMSHYEAMGLSGHKLSPEFDG---KLSGKRLPDLG--LESGDLI	408
Db	488	lspregvavsevmllsgermkvqsgpslklkltstgllklsqkqkrggdeesgeht	547	
QY	409	EV 410		
Db	548	qv 549		
RESULT 11				
ID	AAB41231			
XX	AAB41231	standard; Protein; 2971 AA.		
AC	AAB41231;			
DT	08-FEB-2001	(first entry)		
DE	Human OREF995	polypeptide sequence SEQ ID NO:1990.		
XX	Human	open reading frame; OREF: detection; cytostatic; hepatotropic; vulnerrary; antipsoriatic; antiparkinsonian; neuroprotective; anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antiviral; antibacterial; antifungal; antirheumatic; antithyroid; antianemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosis; infection; severe acquired immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive.		
XX	Homo sapiens.			
OS	Wo200058473-A2.			
PN	05-OCT-2000.			
PD	31-MAR-2000; 2000WO-US08621.			
PF	31-MAR-1999; 99US-0127607.			

PR 30-MAR-2000; 2000US-0540763.
 XX
 XX (CURA-) CURAGEN CORP.
 XX
 PI Shimkets RA, Leach M;
 XX
 XX WPI; 2000-602362/57.
 DR N-PSDB; AAC75440.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 XX Claim 11; Page 1510-1517; 5507pp; English.
 XX
 XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORF open reading frames 1 to 3161. The ORF
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antiproliferative; antiparasitoid; neuroprotective;
 CC osteoplastic; anticonvulsant; antiallergic; immunosuppressive;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antihypertensive; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORF-associated disorder. The
 CC nucleic acids can be used to express ORF proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 CC
 XX
 XX Sequence 2971 AA:
 SQ

Query Match 5.8%; Score 121; DB 21; Length 2971;
 Best Local Similarity 21.2%; Pred. No. 0.53; Indels 166; Gaps 18;
 Matches 102; Conservative 41; Mismatches 173;

6 RGRGPRSGCGARRARG-RC-----PRA 31
 2125 kakperpgtvtserltgaretgahnpvishqtrsttpprcsparearvprpapr 2184
 32 ROSPA---RLIPDTVLVDVSDSEVLVADPVEV-PVARLPAPAKPEQSDSDSEGAA 87
 2185 rptasapaalpavpyv---sapvlsapnpltlpvhllppppsqqlppspact 2241
 88 EGPGAP---RTIVRRRRRLDGEAPVYVSGKQSSNLIPD----- 130
 2242 pppactpppactpppactclvpsppllp---psvpisavtnlpgllpaelcaqa 2298
 131 -----NSSILKICPSE-----PEDEDLTNSGSPS---EDDALP 162
 2299 laspeslelasvasetslslyppkdlpvaivelpysenhlstpsapbltleagship 2358
 163 SGSPWRKKLKKCKEKKME---FPDQDISPLPQSSRN---KSKRHTEALQKL 212
 2359 ngp-----eqeapdsaefttlvlpgeelplcvseanglppsaasdeqp 2408
 213 REVAKRLQDLRSCSPKQHSALQSTDDDEVL-----VEGVLPGSSRLFTLKI 262
 2409 leadrteelkaptspkqgelvtaeapstssatspdpaprr----- 2463
 263 RCRDLVRLPRMSEPLQNVVDHMANHLGVSPNRILLFGESEL-----SPTAPSTLKL 317
 2464 rtssdvelrggtgtrpgpapkylrk-----pdlvtvveekelvyrrrrgrgaastlv 2520
 318 GVADIDCVLASSSEATETSOELRLRVQGEKQHOMLEISLSPDSPKLVLMSHYEAMGL 377

DB 2521 gvs-----etsaspsspsvmsgsspsppl 2546
 QY 378 SG 379
 DB 2547 99 2548

RESULT 12
 ID AAB50363 standard; protein; 2972 AA.
 XX
 AC AAB50363;
 XX
 DT 12-MAR-2001 (first entry)
 XX
 DE Human SRCAP.
 XX
 KW Human; SRCAP; Snf2 related CBP activator protein; antiviral; CREB;
 KW cAMP regulatory element; CREB binding protein; CBP; ATPase;
 KW transcription activation; DEAD box RNA dependent helicase;
 KW adenoviral DBP protein; beta-actin; nuclear receptor; viral infection.
 XX
 OS Homo sapiens.
 XX
 XX W0200073467-A1.
 PN
 PD 07-DEC-2000.
 XX
 XX 25-MAY-2000; 2000WO-US14719.
 PF
 XX 27-MAY-1999; 99US-0136620.
 PR 25-MAY-2000; 2000US-0579181.
 XX
 PA (UYSL-) UNIV SAINT LOUIS.
 XX
 XX Chiviva J, Yaciuk P;
 XX
 DR WPI; 2001-061545/07.
 DR N-PSDB; AAC69860.
 DR
 XX
 PT Snf2 related cAMP regulatory element (CREB) binding protein (CBP)
 PT activator protein, capable of co-activating CREB binding protein,
 PT useful for modulating transcription and for affecting viral infection -
 XX
 PS Claim 1; Page 86-94; 103pp; English.
 XX
 CC The present sequence is an Snf2 related CREB (cAMP regulatory element)
 CC binding protein (CBP) activator protein (SRCAP) polypeptide. It has
 CC ATPase activity and is capable of activating transcription. SRCAP
 CC polypeptides are useful for activating transcription in a cell, for
 CC enhancing CREB (cAMP regulatory element) binding protein (CBP)-mediated
 CC activation of transcription in a cell, for treating a patient having a
 CC disease involving a function such as insufficient transcription of a
 CC gene, e.g. a gene mediated by CBP co-activation, DEAD box RNA dependent
 CC helicase, adenoviral DBP protein, beta-actin or a nuclear receptor
 CC affected by SRCAP protein. Compounds that modulate SRCAP function, such
 CC as antibodies, antisense molecules, polynucleotides or ribozymes, are
 CC useful for treating diseases mediated by SRCAP-activated transcription,
 CC for example, infection by adenovirus, hepatitis C virus, human
 CC immunodeficiency virus type-1, Epstein Barr virus, cytomegalovirus or
 CC hepatitis B virus.
 CC
 XX
 XX Sequence 2972 AA:
 SQ

Query Match 5.8%; Score 121; DB 22; Length 2972;
 Best Local Similarity 21.2%; Pred. No. 0.53;
 Matches 102; Conservative 41; Mismatches 173; Indels 166; Gaps 18;
 6 RGRGPRSGCGARRARG-RC-----PRA 31
 2125 kakperpgtvtserltgaretgahnpvishqtrsttpprcsparearvprpapr 2184

QY	32	ROSFA---	RLIDFTLVLYDVSNSDEEVL	EVADPVEV	-PVARLLPAKAP	PEQDSDS	SEGNA	87											
Db	2185	rlpctasapala	rlpvpvpy---	sapvysa	npmltl	lpvhlips	pppsq	ipccss											
QY	88	EGPACAP-----	-RLVRRRRRL	LDPGEAR	PVPVYSGK	VOSLNL	IPD-----	130											
Db	2242	pppactcp	ppahtrpp	paqctcl	vprsp	lll	ggp---	psvysa											
QY	131	-----	NSSLKLC	PSB-----	PEDEADLT	NSGSS	SPS---	EDDALP											
Db	2299	laspesle	lasvasse	tslsj	vppkdl	lpvave	llpvsek	hls1t											
QY	163	SGSPMR	KRKKECKE	EKKME	-----	FPDDDIS	PLPQ	SSRN---											
Db	2335	ngq-----	-eqeap	saeqct	tlvl	pegel	lplv	senglel											
QY	213	RVNKR	RLDRLR	SCLSPK	HOSP	ALOST	DEVL-----	VEGPVL											
Db	2409	leadtl	seel	teakr	ptsspek	pqgel	vtaev	apars											
QY	263	RCRAD	LVRLP	RMSEPL	QNV	DHMAN	HLGVSP	NRILL											
Db	2464	rt	sadv	elrgg	tyrpg	ppg	pkv	lrlkl---											
QY	318	GVADII	DCV	LV	LV	LV	LV	LV											
Db	2521	gvs-----	-----	-----	-----	-----	-----	-----											
QY	378	SG	379																
Db	2547	gg	2548																
RESULT 13																			
XX	AA	B50362	standard; protein; 3118 AA.																
XX	AC	AA	B50362;																
XX	DT	12-MAR-2001	(first entry)																
XX	DE	Human SRCAP.																	
XX	KW	Human: SRCAP. Snf2 related CBP activator protein; antiviral; CREB; CAMP regulatory element; CREB binding protein; CBP; ATPase; Transcription activation; DEAD box RNA dependent helicase; adenoviral DBP protein; beta-actin; nuclear receptor; viral infection.																	
XX	OS	Homo sapiens.																	
XX	PN	WO200073467-A1.																	
XX	PD	07-DEC-2000.																	
XX	PF	25-MAY-2000; 2000MO-US14719.																	
XX	PR	27-MAY-1999; 99US-0136620.																	
XX	PR	25-MAY-2000; 2000DS-0579181.																	
XX	PA	(UYSL-) UNIV SAINT LOUIS.																	
XX	PI	Chrivia J, Yaciuk P;																	
XX	DR	WPI: 2001-061545/07.																	
XX	XX	N-PSDB: AAC89859.																	
XX	PT	Snf2 related CAMP regulatory element (CREB) binding protein (CBP) activator protein, capable of co-activating CREB binding protein, useful for modulating transcription and for affecting viral infection -																	
XX	PS	Claim 6; Page 77-86; 103pp; English.																	

Query Match	Best Local Similarity	5.8%	Score 121	DB 22	Length 3118
Matches 102	Conservative 41	Mismatches 173	Indels 166	Gaps 18	
6	NGRGRSRSGRGARRARG-	-----RC-----	PRA 31		
Db	2271 kakaperptgrserlrgaetcgahntcvlsabqtrsttprcsparervrprpapr	-----RC-----	PRA 31		
Qy	32 RQSPA--RLIDPTVLVDLSDSDDEVLEAVDPVEV-PAARLPAPAKPPQSDSDEGAA	87			
Db	2331 rfcpsapaalpavlvpvr---sapvysapnpltlpvhllpspppsqippcsspact	2387			
Qy	88 EBPAGAP-----RVLVRRRRRLDPPGAPVPVYSGKVOSSLNLPD-----	130			
Db	2388 pppactpprhthpppqactclvpsaplllgp---psvrlsaevtnlpjldlrpeaelcaga	2444			
Qy	131 -----NSSILKLCPSSE-----PEDEADLTNSGSSPS---EDDALP	162			
Db	2445 laspelslasvasstelsilvpkdlipraveallpvsekhlslpsapsltlleagsrip	2504			
Qy	163 SGGPMRRKLRKCEKEKKME-----PPDDISPLPPSSRN-----KSRKHTMALQKL	212			
Db	2505 nqg-----eqeapdsaeqtlcltvpegeelplcvseengtlelpsaasdepldqp	2554			
Qy	213 REVNKRLODRLRSCLPKQHOQSPALOSTDDEVVL-----VEGVLVQSSRLFTLKI	262			
Db	2555 leadrtseeltakrptsspekprqelvtlaevaaarstssatspspegpsparpprr-----	2609			
Qy	263 RCRARLVLRPVMSEPLQVNDVHMNHGLGVSNRIILLRGESFL-----SPATPSTTLK	317			
Db	2610 rtsadevrlrggtgrpgqpprvlrlk---prglvtveekelvygrrtqrgaaastlirp	2666			
Qy	318 GVAIDIDCVLASSSEATETTSQELRLRVQKKEKHOMLEISLSPDSLKMLSHYEBAMGL	377			
Db	2667 gys-----etsaapgspsvrsmgspspspl	2692			
Qy	378 SG 379				
Db	2693 gq 2694				
RESULT 14	AAW61247				
ID	AAW61247 standard: Protein; 565 AA.				
XX	AAW61247:				
XX	02-OCT-1998 (first entry)				
XX	Streptococcus pneumoniae SP123 protein.				
XX	Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;				
KW	detection; pneumonia; otitis media; meningitis.				

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 11, 2001, 08:48:26 ; Search time 13.53 seconds
(without alignments)
626.994 Million cell updates/sec

Title: US-09-192-611-2

Perfect score: 2099
Sequence: 1 MAEPLRGGRPSRSGRGAR.....GKELPADLGESGLIEVWG 412

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/prodata/2/iaa/5A.COMB.pep.*
2: /cgn2_6/prodata/2/iaa/5B.COMB.pep.*
3: /cgn2_6/prodata/2/iaa/6A.COMB.pep.*
4: /cgn2_6/prodata/2/iaa/6B.COMB.pep.*
5: /cgn2_6/prodata/2/iaa/PCTUS.COMB.pep.*
6: /cgn2_6/prodata/2/iaa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2099	100.0	412	2	US-08-755-584-2
2	2099	100.0	412	3	US-09-192-611-2
3	122.5	5.8	1780	1	US-08-769-309A-5
4	122.5	5.8	1780	3	US-08-994-570-5
5	120.5	5.7	565	4	US-08-961-083-218
6	113.5	5.4	661	2	US-08-795-868-14
7	111	5.3	897	1	US-08-095-737-4
8	111	5.3	897	2	US-08-480-145-4
9	111	5.3	897	2	US-08-477-389-4
10	108.5	5.2	505	1	US-08-631-200-2
11	108.5	5.2	505	1	US-08-630-592-4
12	108.5	5.2	505	1	US-08-714-891-4
13	108.5	5.2	505	2	US-08-829-553-2
14	108.5	5.2	505	2	US-08-922-267A-2
15	108.5	5.2	505	2	US-08-936-707A-2
16	108.5	5.2	505	2	US-08-936-706A-2
17	108.5	5.2	505	3	US-09-248-203-2
18	108.5	5.2	505	3	US-09-032-365A-4
19	108.5	5.2	505	4	US-08-812-824-3
20	108.5	5.2	505	4	US-09-406-071-2
21	108	5.1	459	1	US-08-630-592-2
22	108	5.1	459	1	US-08-714-991-2
23	108	5.1	459	3	US-09-032-365A-2
24	108	5.1	673	2	US-08-455-073A-6
25	108	5.1	1001	4	US-09-060-410-2
26	107	5.1	521	2	US-08-721-684C-2
27	107	5.1	521	2	US-09-005-970-2

28	107	5.1	521	4	US-09-407-715-2	Sequence 2, Appl
29	107	5.1	570	3	US-08-826-246-2	Sequence 2, Appl
30	107	5.1	570	3	US-08-944-495-2	Sequence 2, Appl
31	107	5.1	570	3	US-09-126-640-7	Sequence 7, Appl
32	107	5.1	570	4	US-08-925-588-2	Sequence 2, Appl
33	106.5	5.1	723	2	US-08-548-159-5	Sequence 5, Appl
34	106.5	5.1	1142	2	US-08-993-118-7	Sequence 7, Appl
35	106.5	5.1	1142	3	US-08-845-528C-7	Sequence 7, Appl
36	104.5	5.0	861	1	US-08-484-105-18	Sequence 18, Appl
37	104.5	5.0	861	1	US-08-484-106-18	Sequence 18, Appl
38	104	5.0	126	2	US-08-853-974-1	Sequence 1, Appl
39	104	5.0	126	4	US-08-172-988-1	Sequence 1, Appl
40	104	5.0	3248	1	US-08-353-700-1	Sequence 1, Appl
41	104	5.0	3248	5	PCT-US95-16216-1	Sequence 1, Appl
42	103.5	4.9	586	2	US-08-630-822A-70	Sequence 70, Appl
43	103.5	4.9	586	2	US-09-005-069-70	Sequence 70, Appl
44	103.5	4.9	704	1	US-08-188-582-5	Sequence 5, Appl
45	103.5	4.9	704	1	US-08-646-715-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-08-755-584-2
: Sequence 2, Application US/08755584
: Patent No. 5858711
: GENERAL INFORMATION:
: APPLICANT: Glincher, Laurie H.
: APPLICANT: Hodge, Martin R.
: TITLE OF INVENTION: NF-AT-INTERACTING PROTEIN NIP45 AND METHODS
: TITLE OF INVENTION: OF USE THEREFOR
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: LAHIVE & COCKFIELD
: STREET: 60 State Street, suite 510
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02109-1875
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/755,584
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Giulio A. DeConti, Jr.
: REGISTRATION NUMBER: 31,503
: REFERENCE/DOCKET NUMBER: HUI-026
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617)227-7400
: TELEFAX: (617)227-5941
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 412 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-755-584-2

Query Match 100.0%; Score 2099; DB 2; Length 412;
Best Local Similarity 100.0%; Pred. No. 1 6e-174;
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MAEPLRGGRPSRSGRGARARAGRGRCPRARQSPARLIPDTVLVDLVSDSEVLEVAD 60

|||||
Db 1 MAEPLRGGRPRSGRGARRAGARCGRCPRAROSPARIIDPTVYLVDSDEVELEVAD 60
QY 61 PVEVPVAPRLAPAKPPEDSDSDSGAEGPAGAPRTLYRRRRRLDPGEPVVPVYSGK 120
|||||
Db 61 PVEVPVAPRLAPAKPPEDSDSDSGAEGPAGAPRTLYRRRRRLDPGEPVVPVYSGK 120
QY 121 VQSSNLIPNSSLKLCPSPEDEADLTNSGSSPSDDALPSGSPWRKLRKCKEKEE 180
|||||
Db 121 VQSSNLIPNSSLKLCPSPEDEADLTNSGSSPSDDALPSGSPWRKLRKCKEKEE 180
QY 121 VQSSNLIPNSSLKLCPSPEDEADLTNSGSSPSDDALPSGSPWRKLRKCKEKEE 180
181 KMEFPDQDISPLPQPSRRKSRKHTALQKLRVNRRLDLSCLSPKQHSFALQSTD 240
Db 181 KMEFPDQDISPLPQPSRRKSRKHTALQKLRVNRRLDLSCLSPKQHSFALQSTD 240
QY 241 DEVYLVGAPVLPOSSRLFTLKIRCRADLVRLPVMSPEPLQNVVDHMANHGVSPNRITLL 300
Db 241 DEVYLVGAPVLPOSSRLFTLKIRCRADLVRLPVMSPEPLQNVVDHMANHGVSPNRITLL 300
QY 301 FGESELSPTATPSTLKLGVAIDIDCVVLASSSEATETSQELRLRVQGEKQHMLEISLSP 360
Db 301 FGESELSPTATPSTLKLGVAIDIDCVVLASSSEATETSQELRLRVQGEKQHMLEISLSP 360
QY 361 DSPKLVMSHYEAMGSLGKLSFFFDGTRKLSGKELPADLGESGDIIEWWG 412
Db 361 DSPKLVMSHYEAMGSLGKLSFFFDGTRKLSGKELPADLGESGDIIEWWG 412

RESULT 2

US-09-192-611-2
; Sequence 2, Application US/09192611
; Patent No. 6090561
; GENERAL INFORMATION:
; APPLICANT: Glimcher, Laurie H.
; TITLE OF INVENTION: NF-AT-INTERACTING PROTEIN NIP45 AND METHODS
; TITLE OF INVENTION: OF USE THEREFOR
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/192,611
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/755,584
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Giulio A. DeConti, Jr.
; REGISTRATION NUMBER: 31,503
; REFERENCE/DOCKET NUMBER: HUT-026
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 412 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-192-611-2

Query Match 100.0%; Score 2099; DB 3; Length 412;
Best Local Similarity 100.0%; Pred. No. 1,6e-174;
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEPLRGGRPRSGRGARRAGARCGRCPRAROSPARIIDPTVYLVDSDEVELEVAD 60
Db 1 MAEPLRGGRPRSGRGARRAGARCGRCPRAROSPARIIDPTVYLVDSDEVELEVAD 60
QY 61 PVEVPVAPRLAPAKPPEDSDSDSGAEGPAGAPRTLYRRRRRLDPGEPVVPVYSGK 120
Db 61 PVEVPVAPRLAPAKPPEDSDSDSGAEGPAGAPRTLYRRRRRLDPGEPVVPVYSGK 120
QY 121 VQSSNLIPNSSLKLCPSPEDEADLTNSGSSPSDDALPSGSPWRKLRKCKEKEE 180
Db 121 VQSSNLIPNSSLKLCPSPEDEADLTNSGSSPSDDALPSGSPWRKLRKCKEKEE 180
QY 121 VQSSNLIPNSSLKLCPSPEDEADLTNSGSSPSDDALPSGSPWRKLRKCKEKEE 180
181 KMEFPDQDISPLPQPSRRKSRKHTALQKLRVNRRLDLSCLSPKQHSFALQSTD 240
Db 181 KMEFPDQDISPLPQPSRRKSRKHTALQKLRVNRRLDLSCLSPKQHSFALQSTD 240
QY 241 DEVYLVGAPVLPOSSRLFTLKIRCRADLVRLPVMSPEPLQNVVDHMANHGVSPNRITLL 300
Db 241 DEVYLVGAPVLPOSSRLFTLKIRCRADLVRLPVMSPEPLQNVVDHMANHGVSPNRITLL 300
QY 301 FGESELSPTATPSTLKLGVAIDIDCVVLASSSEATETSQELRLRVQGEKQHMLEISLSP 360
Db 301 FGESELSPTATPSTLKLGVAIDIDCVVLASSSEATETSQELRLRVQGEKQHMLEISLSP 360
QY 361 DSPKLVMSHYEAMGSLGKLSFFFDGTRKLSGKELPADLGESGDIIEWWG 412
Db 361 DSPKLVMSHYEAMGSLGKLSFFFDGTRKLSGKELPADLGESGDIIEWWG 412

RESULT 3

US-08-769-309A-5
; Sequence 5, Application US/08769309A
; Patent No. 5741890
; GENERAL INFORMATION:
; APPLICANT: Scott, John D.,
; APPLICANT: Nauert, Brian J.,
; APPLICANT: Klauert, Theresa M.
; TITLE OF INVENTION: Protein Binding Domains of Gravin
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower/233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/769,309A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5741890and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/33451
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1780 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein


```

; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 218:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 565 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-218

Query Match      5.7%; Score 120.5; DB 4; Length 565;
Best Local Similarity 20.9%; Pred. No. 0.019;
Matches 81; Conservative 66; Mismatches 156; Indels 85; Gaps 19;

QY 29 PRAROSP-----ATLIDTVLYVSDSEVLEVADPVEV--PVARLP--AAKPE-Q 77
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
Db 34 PKEESKEPKSEVKPTDPTLPRKVEGKDSAPAPVEEVEGSEKPEEKAVKPEEQ 93
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
QY 78 DSDSDSGAAEGPAGARTLVRRRRRLDPGEAPVVPVYSGKVOSS---LNLIPDNSSL 134
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
Db 94 PSDBAPAESVEQAGEP-VAPREDEKAPVEKEKPEAPEEKAVEETPKQEESTPDTRAE 152
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
QY 135 LKICP-----SEPDADLTNNGS--SPSDDALPSSGPPRRKK 170
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
Db 153 EYEPKKEEYFNOSIEQPKVETPAVEKQTEPEEKVEQAGEPAPRDEQAPT-APEPE 211
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
QY 171 LRRKCEKEKMEEPKPPDISPLPQPSRKMSRKHTFALQKLRVNNRLDRLRCLSPKQ 230
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
Db 212 KQEVPEEKAVEE-----TPKPEDKIKIGTKERPVDE-SELNNOI-DKASSVSPTD 261
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
QY 231 HOSFALOSTDEEVLYGEPVLPQSSRLFTLKIRGADLVRLPVMSSE--PLQNVVD--- 284
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
Db 262 YSTASYAL-----GPVLETKAGVYA-----SEPVKQPEVNSGSETKTKTIDALNV 307
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
QY 285 ---HMANHLGVSPNRILLFGESELSPTAPPSTLKLGVADIIDCVVLASSEATSEATSOEL 341
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
Db 308 DKTELNTIADAKTKVKEHSDRSQNMLOTEVTKAEKVAANTD---AKQSEVNAVEEKL 363
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
QY 342 RLRYQKEKHQMLEISLSPSPKLVMS 369
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
Db 364 YATLE-----KLVELS---EKPIITLTS 383
   | : || : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
US-08-795-868-14
; Sequence 14, Application US/08795868
; Patent No. 5846773
; GENERAL INFORMATION:
; APPLICANT: Lee, Mu-En
; APPLICANT: Hsieh, Chung-Ming
; TITLE OF INVENTION: A SINGLE GENE ENCODING AORTIC-SPECIFIC
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P. C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,868
; FILING DATE: 06-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/494,577
; FILING DATE: 22-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 05433/032001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 661 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-08-795-868-14

Query Match      5.4%; Score 113.5; DB 2; Length 661;
Best Local Similarity 20.0%; Pred. No. 0.096;
Matches 103; Conservative 55; Mismatches 131; Indels 227; Gaps 25;

QY 4 PLRGR---GPRSRG---RGARRRGAR----- 25
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
Db 26 PARGHPIVAPRSHSGACAGAHHRREAREVYVVALPGHRESRPQTPLSAAGRLMA 85
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
QY 26 -GRCPRAROSPARIPTVLYVSDSE--EYLEVADPVEVPV-----ARLPAPAK 74
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
Db 86 LGSNRLVRRAGSRL-----DKLOFEERRSLIENSPPAPLRPVPLRKASLEQPK 139
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
QY 75 PEODSDSDSGAAEGPAGARTLVRRRR-----RLDPGEAPVVPVY-----GK 120
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
Db 140 SERGAPWGTGASGSELRAFGASVARRRLFOQKAASLDERTRQSPASDLELRFAQLGR 199
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
QY 121 VOSSLN---LIPDNSSL---LKICPSEP--DEADLTNNGSSPSDDAL----- 161
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
Db 200 IRRSTSEELVRSHEISRLATLQRAPSPREPGEPEPLFSRPSTPSTRAYSMAAQPSPS 259
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
QY 162 ---PSSGPPMKLRRKC-----EKERKMEEP-----DDIISLPQPS----- 197
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
Db 260 AEKPGDEGRPRSRGPAGRTPEGEGPOQEVRRRDQFPLTRSRAIOECRSVPPEADPE 319
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
QY 198 SRNKS---RKHTFALQKLR-----EVNKRLO----- 220
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
Db 320 ARTKAPRGKRREPPAQAVRLPMATPGLGAAVPTLEKNRAGPEAKRLRRGPEEDGPW 379
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
QY 221 ---DLRSCLSPKQHQ-----SPALQSTDEEVL-----VEPVLPQSSRLFTLKIRCRAD 267
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
Db 380 GPMDBRGARQOGKRRAPRTPSELESSDSDSVSAGEEPLAPV----- 422
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
QY 268 LVRLPVMSSEPLQNVYDHMANHLGVSPNRILLFGESELSPTATPSTLKLGVADIIDCVV 327
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
Db 423 -----FELPLQNVV-----VAP-----GADVLLACTI 444
   | : || : : : : : : : : : : : : : : : : : : : : : : : :

QY 328 LAS-----SSEATETSOELRLRYQKEKHQML 354
   | : || : : : : : : : : : : : : : : : : : : : : : : : :
Db 445 TANPPQVSWHKGSALRSRGRLLRREG-ERRITLL 479
   | : || : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
US-08-095-737-4
; Sequence 4, Application US/08095737
; Patent No. 5487979
; GENERAL INFORMATION:
; APPLICANT: Difiore, Pier P
; APPLICANT: FazioII, Francesca

```


Db 541 HAEGONNLESEPTHOESSVSSPEIAPSDV--TDESAVTVAGNEKYTPRFDDKHSKEE 598
QY 396 LPADLGESGDLE 409
Db 599 DP--FNWESSLTD 610

RESULT 9

US-08-477-389-4
; Sequence 4, Application US/08477389
; Patent No. 5872219
; GENERAL INFORMATION:
; APPLICANT: DiFiore, Pier P
; APPLICANT: Fazioli, Francesca
; TITLE OF INVENTION: A Substrate for the Epidermal Growth
; FACTOR RECEPTOR KINASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: United States of America
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,389
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/095,737
; FILING DATE: 22-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel, Ned A
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH60,001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 897 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-477-389-4

Query Match 5.3%; Score 111; DB 2; Length 897;
Best Local Similarity 21.2%; Pred. No. 0.24;
Matches 92; Conservative 79; Mismatches 153; Indels 110; Gaps 24;

QY 61 PVEVVARLPL-----APA-KPEOD-----SDSDSEGAAG-----PAGAPRTL 97
Db 202 PMSIPPALVPPSKRTWTVAPEAKATDELFTLTKDMDYVSGLEVREFFLTGLPSAL 261
QY 98 VRRRRRLDGPGEAPVVPVYS-----GKVOS-----SLNLIPDNSSLK-----LC 138
Db 262 LAH-----INSLCDTKGCGKLSKDQFALAHFLI--NQKLKIGIDPPHSLT 304
QY 139 PS--EPDEADLTN--SGSSPSED-----DALPSSPPRKRLRKCKEKKMEEPD 187
Db 305 PEMIPSDRSLSOKNITGSSPVADFSAIKELDTLNNEIIVDLQREKKNVEODLKEKEEDTVK 364
QY 188 QDISPLPQPSRKNRSRKTALQKLRVNRKRLDRLRSCLSPKQHQ-----SPALQSTNDE 242
Db 365 QRTSEV--QDLQDEVOKESTINQLQAKQKQVOVDLLEGLDEQKALQLEQLOAEVRRKCAEE 422

QY 243 VLV---EGPVLPOSSRLFTLK---IRCRADLVRLVRMSEPLQNV-----VDHMANHL 290
Db 423 AOLISLKAEITQESQISSEYEEELKARELSRLQOETRAQLEESVSGAQLPEPQOHL 482
QY 291 GVSPNNILLFGESELSPTAT-----PSTKLGVADITIDCVVLASSSEATETSO 339
Db 483 QESQOETISSQMRLKEMKDLTDNNQSNWSSPQSVLVNGATDY--CSLSSTSSSETNFMKE 540
QY 340 --ELRLRVQGEKHKOMLEISPD--SPKVLMSHYEAMQLSGH--KLSPFDQTKLSGKE 395
Db 541 HAEGONNLESEPTHOESSVSSPEIAPSDV--TDESAVTVAGNEKYTPRFDDKHSKEE 598
QY 396 LPADLGESGDLE 409
Db 599 DP--FNWESSLTD 610

RESULT 10

US-08-631-200-2
; Sequence 2, Application US/08631200
; Patent No. 5646040
; GENERAL INFORMATION:
; APPLICANT: Kleyn, Patrick W.
; APPLICANT: Moore, Karen J.
; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
; DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/631,200
; FILING DATE: 12-APR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-057
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 505 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-631-200-2

Query Match 5.2%; Score 108.5; DB 1; Length 505;
Best Local Similarity 22.9%; Pred. No. 0.18;
Matches 64; Conservative 36; Mismatches 114; Indels 65; Gaps 10;

QY 3 EPLR-----GRGPRSRGGGARARAGARCRPARQSPARLIPDVIVLDVSDSEEV 56
Db 44 EPLMVQANADGR--PRSR-----RAQSEDA--PLVESYSSSGSTSY 83
QY 57 EVADPVEVVARLPAPAKPRQSDSDSEGAEE--GPAGAPRTLVRRRRRLDGPGEAPV 114
Db 84 QVEDADSIASVOGATRRPARPAKSKSGAAGGOGGAPRKKKKHKTGSPFA----- 138
QY 115 PVYSGKVQSSSLNLIPDNS-----SLKLCPSPEDEADLTNSGSSPSDDALPSSGPPWR 168

Db 139 -----TLAEDKSEAQGVQILTVGOSDHDKDGETAAGG-----AOPSGODLR 182
QY 169 KKLKKCKEKKKEEPPDDDISPLPQSSRNKSRKTEALQKIREVKNKRLQDLRSLSP 228
Db 183 ATMQRKGISSMSFDEDEDED-----ENSSSSQLNSNTRPPSSATSATRSKSIREAASAPSP 236
QY 229 KOHOSPALOSTDDEVVLVEGVPVLPQSSRLFTLKIRCRAD 267
Db 237 AAPEPP-----VDIEVODLEEFALRPAPQGITIKCRITRD 271

RESULT 11
US-08-630-592-4
; Sequence 4, Application US/08630592
; Patent No. 5770432
; GENERAL INFORMATION:
; APPLICANT: Nishina, Patsy
; APPLICANT: No. 5770432entrauth, Konrad
; APPLICANT: Nagert, Juergen
; APPLICANT: No. 5770432th, Michael
; TITLE OF INVENTION: Obesity Associated Genes
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 3400 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 941114187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MSDOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630, 592
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sherwood, Pamela J.
; REGISTRATION NUMBER: 36,677
; REFERENCE/DOCKET NUMBER: A59504/BIR/PJS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 7811989
; TELEFAX: (415) 3983249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 505 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-630-592-4

Query Match 5.2%; Score 108.5; DB 1; Length 505;
Best Local Similarity 22.9%; Pred. No. 0.18;
Matches 64; Conservative 36; Mismatches 114; Indels 65; Gaps 10;

QY 3 EPLR-----GRGPRSRGCGARRARGRCRRAROSPRLIPDYLVLVDSDEVL 56
Db 44 EPLMVQANADGR-PRSR-----RARSERDA---PLVESTYSSGSTSY 83
QY 57 EVADPVEVPYARLPAPKPEODSDSEGAEE--GPAAGPRTIVRRRRRLLDGEAPV 114
Db 84 QVQEADSIASVOLGATPPAPASAKKSGAASGOGGAPRKKEKKGHKGTSGA-----138
QY 115 PYVSGKVOSSLNLPDMS-----SLKLCPSEPEDEADLTNSGSSPEDDALPSSGSPWR 168
Db 139 -----TLAEDKSEAQGVQILTVGOSDHDKDGETAAGG-----AOPSGODLR 182

QY 169 KKLKKCKEKKKEEPPDDDISPLPQSSRNKSRKTEALQKIREVKNKRLQDLRSLSP 228
Db 183 ATMQRKGISSMSFDEDEDED-----ENSSSSQLNSNTRPPSSATSATRSKSIREAASAPSP 236
QY 229 KOHOSPALOSTDDEVVLVEGVPVLPQSSRLFTLKIRCRAD 267
Db 237 AAPEPP-----VDIEVODLEEFALRPAPQGITIKCRITRD 271

RESULT 12
US-08-714-991-4
; Sequence 4, Application US/08714991
; Patent No. 5776762
; GENERAL INFORMATION:
; APPLICANT: NORTH, Michael
; APPLICANT: NISHINA, Patsy
; APPLICANT: No. 5776762en-Trauth, Konrad
; APPLICANT: Nagert, Juergen
; TITLE OF INVENTION: OBESITY ASSOCIATED GENES
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/714, 991
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SHERWOOD, Pamela J.
; REGISTRATION NUMBER: 36,677
; REFERENCE/DOCKET NUMBER: A-59504-1/PJS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-494-8771
; TELEFAX: 415-494-8700
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 505 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-714-991-4

Query Match 5.2%; Score 108.5; DB 1; Length 505;
Best Local Similarity 22.9%; Pred. No. 0.18;
Matches 64; Conservative 36; Mismatches 114; Indels 65; Gaps 10;

QY 3 EPLR-----GRGPRSRGCGARRARGRCRRAROSPRLIPDYLVLVDSDEVL 56
Db 44 EPLMVQANADGR-PRSR-----RARSERDA---PLVESTYSSGSTSY 83
QY 57 EVADPVEVPYARLPAPKPEODSDSEGAEE--GPAAGPRTIVRRRRRLLDGEAPV 114
Db 84 QVQEADSIASVOLGATPPAPASAKKSGAASGOGGAPRKKEKKGHKGTSGA-----138
QY 115 PYVSGKVOSSLNLPDMS-----SLKLCPSEPEDEADLTNSGSSPEDDALPSSGSPWR 168
Db 139 -----TLAEDKSEAQGVQILTVGOSDHDKDGETAAGG-----AOPSGODLR 182
QY 169 KKLKKCKEKKKEEPPDDDISPLPQSSRNKSRKTEALQKIREVKNKRLQDLRSLSP 228
Db 183 ATMQRKGISSMSFDEDEDED-----ENSSSSQLNSNTRPPSSATSATRSKSIREAASAPSP 236

QY 229 KQHSPALQSTDEVLVVEGPVLPOSSRLFTLKIRCRAD 267
: | | | | : : : : :
Db 237 AAEPP-----VDIEVODLEEFALRPAPOGITIKCRITRD 271

RESULT 13
US-08-829-553-2
; Sequence 2, Application US/08829553
; Patent No. 5817762
; GENERAL INFORMATION:
; APPLICANT: Kieyn, Patrick W.
; APPLICANT: Moore, Karen J.
; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennle & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/829,553
; FILING DATE: 28-MAR-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/631,200
; FILING DATE: 12-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-057
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNTE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 505 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-829-553-2

Query Match 5.2%; Score 108.5; DB 2; Length 505;
Best Local Similarity 22.9%; Pred. No. 0.18;
Matches 64; Conservative 36; Mismatches 114; Indels 65; Gaps 10;
QY 3 EPLR-----GRGPRSRGGGARRARGRCPRARQSPARLLPDIYLVIVDSDSDEVL 56
: | | | | : : : : :
Db 44 EPLMVQANADGR-PRSR-----RARQSEBOA---PLVESYLLSSGSTSY 83
: | | | | : : : : :
QY 57 EVADPVEVPVAPARLPAPAKPPRODSDSSEGAEE--GPAGARPTLVRRRRRLDPGEAPV 114
: | | | | : : : : :
Db 84 QVEADSIASVQIGATRPAPASAKSKGAASGGGAGPRKKKHKGTSGPA----- 138
: | | | | : : : : :
QY 115 PYVSGKVQSSNLIPDNS-----SLKLCPSPEDEADLTNNGSSPSSEDDALPSGSPWR 168
: | | | | : : : : :
Db 139 -----TLADKSEAGVPVQILITVGQSDHDKDAGETAAGC-----AOPSGODLR 182
: | | | | : : : : :
QY 169 KLRKCKEKKKKKEEPPDDISPLPQSSSRNKSRRKHTALQKLRVYKRLQDLRSCISP 228
: | | | | : : : : :
Db 183 ATMQRKGISSMSFDEDEDED-----ENSSSSQULNSNTRPSSATSRKKSIREAASAPSP 236
: | | | | : : : : :
QY 229 KQHSPALQSTDEVLVVEGPVLPOSSRLFTLKIRCRAD 267

Db 237 AAEPP-----VDIEVODLEEFALRPAPOGITIKCRITRD 271
: | | | | : : : : :
Db 237 AAEPP-----VDIEVODLEEFALRPAPOGITIKCRITRD 271

RESULT 14
US-08-922-267A-2
; Sequence 2, Application US/08922267A
; Patent No. 5861239
; GENERAL INFORMATION:
; APPLICANT: Kieyn, Patrick W.
; APPLICANT: Moore, Karen J.
; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennle & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/922,267A
; FILING DATE: 2-SEP-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/829,553
; FILING DATE: 28-MAR-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/631,200
; FILING DATE: 12-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-085
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNTE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 505 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-922-267A-2

Query Match 5.2%; Score 108.5; DB 2; Length 505;
Best Local Similarity 22.9%; Pred. No. 0.18;
Matches 64; Conservative 36; Mismatches 114; Indels 65; Gaps 10;
QY 3 EPLR-----GRGPRSRGGGARRARGRCPRARQSPARLLPDIYLVIVDSDSDEVL 56
: | | | | : : : : :
Db 44 EPLMVQANADGR-PRSR-----RARQSEBOA---PLVESYLLSSGSTSY 83
: | | | | : : : : :
QY 57 EVADPVEVPVAPARLPAPAKPPRODSDSSEGAEE--GPAGARPTLVRRRRRLDPGEAPV 114
: | | | | : : : : :
Db 84 QVEADSIASVQIGATRPAPASAKSKGAASGGGAGPRKKKHKGTSGPA----- 138
: | | | | : : : : :
QY 115 PYVSGKVQSSNLIPDNS-----SLKLCPSPEDEADLTNNGSSPSSEDDALPSGSPWR 168
: | | | | : : : : :
Db 139 -----TLADKSEAGVPVQILITVGQSDHDKDAGETAAGC-----AOPSGODLR 182
: | | | | : : : : :
QY 169 KLRKCKEKKKKKEEPPDDISPLPQSSSRNKSRRKHTALQKLRVYKRLQDLRSCISP 228
: | | | | : : : : :
QY 229 KQHSPALQSTDEVLVVEGPVLPOSSRLFTLKIRCRAD 267

Db 183 ATMORKGISMSRDEDEDED-----ENSSSSQLNSNTRPSSATSRKSTIREAASAPSP 236

QY 229 KOHOSPALQSTDEVVLVEGFLVLPQSSRLFTLKIRCRAD 267

Db 237 AAPEPP-----VDIEVQDLEEFALRPAPGKITKCRITRD 271

RESULT 15

US-08-936-707A-2

; Sequence 2, Application US/08936707A

; Patent No. 5871931

; GENERAL INFORMATION:

; APPLICANT: Kleyu, Patrick W.

; APPLICANT: Moore, Karen J.

; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND

; TITLE OF INVENTION: DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY

; NUMBER OF SEQUENCES: 60

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds LLP

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036-2711

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/936.707A

; FILING DATE: 24-SEP-1997

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: Cortuzzi, Laura A.

; REGISTRATION NUMBER: 30,742

; REFERENCE/DOCKET NUMBER: 7853-100

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 790-9090

; TELEFAX: (212) 869-9741/8864

; TELETYPE: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 505 amino acids

; TYPE: amino acid

; TOPOLOGY: unknown

; MOLECULE TYPE: protein

; US-08-936-707A-2

Query Match 5.2%; Score 108.5; DB 2; Length 505;

Best Local Similarity 22.9%; Pred. No. 0.18; Mismatches 114; Indels 65; Gaps 10;

Matches 64; Conservative 36; Mismatches 114; Indels 65; Gaps 10;

QY 3 EPLR-----GRGPRSRGRCGARARAGRCPPAROSPAPRLIPDVLVDSDEEVL 56

Db 44 EPLAVQANADGR-DRSR-----RARGSEQA---PLVESTYSSSGSTSY 83

QY 57 EVADPVEVPVAKRLPAKPEQSDSDSEGAEE--GPAGAPFTLVRRRRRLIDPEARVY 114

Db 84 QVOEADSTASYVOLGATRPAPASAKSKGAAGSGGAPRKEKKGKHKGTSGPA---- 138

QY 115 PYVSGKVOSSLNLIPDNS-----SLKLCPSEPEDEADLTNSGSSPSEDALPSGSPWR 168

Db 139 -----TLAEDKSEAGPVOILTVGSDHDKDAGETAAGG-----AOPSGODLR 182

QY 169 KLLRKCKEKEKKMEPEPDODISPLPOPSSRNKSKRKHTEALQKLEVNKRLODLRCLSP 228

Db 183 ATMORKGISMSRDEDEDED-----ENSSSSQLNSNTRPSSATSRKSIREAASAPSP 236

QY 229 KOHOSPALQSTDEVVLVEGFLVLPQSSRLFTLKIRCRAD 267

Db 237 AAPEPP-----VDIEVQDLEEFALRPAPGKITKCRITRD 271

Search completed: September 11, 2001, 08:48:55
Job time: 29 sec

